

	APPLICANT: Kronmal, Gregory S.	
	APPLICANT: Lauer, Peter M.	
	APPLICANT: Ruddy, David A.	
	APPLICANT: Thomas, Winston	
	APPLICANT: Tsuchihashi, Zenta	
	APPLICANT: Wolff, Roger K.	
	TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el	
	TITLE OF INVENTION: Sequences and Antibodies Thereeto	
	NUMBER OF SEQUENCES: 31	
	CORRESPONDENCE ADDRESS:	
	ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP	
	STREET: Two Embarcadero Center, 8th Floor	
	CITY: San Francisco	
	STATE: CA	
	COUNTRY: USA	
	ZIP: 94111-3834	
	COMPUTER READABLE FORM:	
	MEDIUM TYPE: Floppy disk	
	OPERATING SYSTEM: PC-DOS/MS-DOS	
	SOFTWARE: PatentIn Release #1.0, Version #1.30	
	CURRENT APPLICATION DATA:	
	APPLICATION NUMBER: US/08/724,394A	
	FILING DATE: 01-OCT-1996	
	CLASSIFICATION: 536	
	ATTORNEY/AGENT INFORMATION:	
	NAME: Fitts, Renee A.	
	REGISTRATION NUMBER: 35,136	
	REFERENCE/DOCKET NUMBER: 017957-000100	
	TELEPHONE: 415-576-0200	
	TELEFAX: 415-576-0300	
	INFORMATION FOR SEQ ID NO: 21:	
	SEQUENCE CHARACTERISTICS:	
	LENGTH: 246240 base pairs	
	TYPE: nucleic acid	
	STRANDEDNESS: not relevant	
	TOPOLOGY: not relevant	
	MOLECULE TYPE: cDNA	
	FEATURE:	
	NAME/KEY: misc.feature	
	LOCATION: 1..246240	
	OTHER INFORMATION: /note= "HLA-H.CONTIG"	
	US-08-724-394A-20	
	Query Match 3.6%; Score 199.2; DB 2; Length 246240;	
	Best Local Similarity 61.7%; Pred.No.2.9e-32;	
	Matches 439; Conservative 0; Mismatches 233; Indels 40; Gaps 6;	
QY	3199 TTTTATGATTCATTTTTATGTTTTTGATGTTTTAAAAAGACAATCTCACCTC 3258	
DB		
	3311 TTTTATAAGACTTTTTTTTTTTTTTTTTTTTTTTTTTTTGACGGAGTCTGCTCT 3252	
QY	3259 GTTACCAGCGGTGAGTGCAGTGGTCCATCATAGCTTCTCGAGTCTGAAGTCTGGG 3318	
DB		
	3251 GTGCCAGCGGTGAGTGCAGTGGCGGGATCTCGGCTCAAGCTCCGGCTCCGGG 3192	
QY	3319 CTCAAGCAATCCCTGCTGCGCTGCCCTCCCAAAGTGTGGGA-TACAGTCATGAGCCACTG 3377	
DB		
	3191 TTCACGCCATTCCTTGCTGCTCAGCTCCCAAGTAGCTGGAGTACAGGCGCCGCCACTA 3132	
QY	3378 CATCTGGC--CTAGGATCCATTTAGATTAAATATATGATTTAAAATTAATATG 3435	
DB		
	3131 CGCCCGGCTAAATTTTGTGATTTTGTAGTAGACAGGGGTTTACCGCTTTAGCCGGGATG 3072	
QY	3436 GCTAATTTT-----ACCTATGTAATGTGATATCT 3466	
DB		
	3071 GTCTGATCTCTGACCTGCTGATCGCGCGCTCGGCTCCCAAAGTGTGGGATTACA 3012	
QY	3467 GGTAAATAATCTAGTTTGTGCTGCTAAAGTTAAAGTGCCTTCCAAATAGCTCATGACG 3526	
DB		
	3011 GGCGTAGGACCGCGCGCGGCTTATAAGACTTTTTAGATAAAACAGAATCTATATC 2952	

QY	3527 TGAGGGGAGACATTTAAAGTGAACAGACAGCAGCGGTGGTGGCTTCAGCCCTGTTAATCC 3588
DB	
	2951 TCCTAATGTATGTTAAGGATGAACC-CGGGCTGGCGCAGTGGCTCACACCTGTTAATCC 2893
QY	3587 CAGCACTCTGGGAGGCTGAGTGGGTGGATCCCTTCAGCCCCTGAGAGTTCGAAGACCAAGCCT 3646
DB	
	2892 CAGCACTTTGGGAGGCGGAGCGGTGAGATCACTTGAGGTTCAGGAGTTTCGAGACCAAGCCT 2833
QY	3647 GAGCAACATGGCAAAACCCCTGTTTCT-----ATAACAAAAAATTAGCCGGCATGTGGC 3700
DB	
	2832 GCCAACACGCGCAAAACCCCTGCTCTACTAAAAATACAAAAATTAGCTGGCATGTGGC 2773
QY	3701 ATGTGCTCTGTGCCAGCTACTAGG-GGCTGAGGACGAGAAATCTTTTGAGACCCAGGA 3759
DB	
	2772 AGGCACCTGTAATCCAGCTGCTTGGGAGACTGAGGACGAGAAATCGTTGAACCTGGGA 2713
QY	3760 GGTCAAGGCTGCAGCTGAGCAGTCTTGGCCCCACTGCACCTCCAGCCTGGGCGACAGGACCA 3819
DB	
	2712 CGTGGGGATTGCAGTGAGCCAAGACTGCACCTCCAGCCTGGGCGACAGGCTGA 2653
QY	3820 GACCTTGCTCTCAAAAAAATAAGAAGAAAAATTTAAAAATAAATGGAACCACT 3871
DB	
	2652 GACACGCTCTCAAAAAATAAAAATAAAAATAAAAATGAACCCCTCT 2601

RESULT 6
US-08-724-394A-21/c
Sequence 21, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Kronmal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Thereeto
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-20

; OTHER INFORMATION: /note= "HLA-H. CONTIG"
US-08-724-394A-21

	Query Match	3.6%	Score 199.2;	DB 2;	Length 246240;
	Best Local Similarity	61.7%;	Pred. No. 2.9e-32;		
	Matches 439;	Conservative 0;	Mismatches 233;	Indels 40;	Gaps 6;
QY	3199	TTTTATGATTCACTTATGTCCTTTTATGTTTTTGATTTTTTAAAGACAAATCTCACCCCT	3258		
Db	3311	TTTTATAAAGACTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGAGCGAGTCTGTGCTCT	3252		
QY	3259	GTTACCAGCGCTGGAGTCAGTGGTGCACATCATAGCTTTCTGCAGTCTTGAACCTCCTGGG	3318		
Db	3251	GTGCCCCAGCGCTGGAGTCAGTGGCGGGATCTCGGCTCACTGCAGAAGCTCCGCCCTCCGGG	3192		
QY	3319	CTCAAGCAATCTCTCCGCCCTTGGCCCTCCCAAAGTGTGGGA-TACAGTCATGAGCCACTG	3377		
Db	3191	TTTCAGCCCAATCTCTGCTCCAGCTCCCAAGTAGCTGGAGCTACAGGCCGCCGCCACTA	3132		
QY	3378	CATCTGGC--CTPAGGATCCATTAGATTAAAATATGCATTTTAAATTTTAAAAATATATG	3435		
Db	3131	CGCCCGGCTPAATTTTGTGATTTTGTAGTAGAGACGGGTTTCCACCGTTTATAGCCGGATG	3072		
QY	3436	GCTAATTTTT-----ACCTTATGTAATCTGTATACT	3466		
Db	3071	GTCTCGATCTCTGACCTCGTATCGGCCCGCTCGGCCTCCAAAGTCTGGGATTACA	3012		
QY	3467	GCTAATAAATCTAGTTTGTGCTTAAAGTTTAAAGTGTCTTCCAATAAGCTTTCATGTAGG	3526		
Db	3011	GGCGTGAACCCACCGCGCCCGCTTATAAAGACTTTTTTAGATAAAAACAGAAATCTATATC	2952		
QY	3527	TCAGGGGAGACATTTTAAAGTGAACAGACAGCCAGGTGTGGTGCCTCAGCCCTGTAATCC	3586		
Db	2951	TCTAATGTATGTATTAGAGATGAACC-CGGCTGGCGCAGTGGCTCACACCTGTAATCC	2893		
QY	3587	CAGCACTCTGGAGBCTGAGTGGGTGGATCGCTTGAGCCCTGGAGTTTCAAGACCAGCT	3646		
Db	2892	CAGCACTTTGGAGGCCGAGCGGGTATCATCTGAGGTCAAGAGTTTCAGAACCCAGCT	2833		
QY	3647	GAGCAACATGCAAAACCCCTGTTC-----ATAACAAAATTAGCCGGCATGGTGGC	3700		
Db	2832	GGCAACACGGCAACCCCTCTCTACTAAAAATACAAAATTAGCTGGCATGGTGGC	2773		
QY	3701	ATGTGCTGTGGTCCCAGCTACTAGG-GGCTGAGGCAGGAGAAATCTTTGAGAGCCAGGA	3759		
Db	2772	AGGCACCTGTAATCCAGCTGCTTTGGAGACTGAGGCAGGAGAAATCGTTTGAACCTGGGA	2713		
QY	3760	GGTCAAGGCTGCAGTGAAGTGTCTGGCCACTGCACTCCAGGCTGGGTACAGGACCA	3819		
Db	2712	CGTGGGATTGCACTGAGCCAAAGCTGCACCACCTGCATCCAGCCTGGGCACAGGGTGA	2653		
QY	3820	GACCTTGCCTCAAAAAATRAGAGAAAAATTTAAAAATTAATGGAACAACACT	3871		
Db	2652	GACACGCTCAAAAAATAAAAAATAAAATAAAAAATAAAAAATGACCCCTCT	2601		

RESULT 7

US-08-724-394A-22/c
; Sequence 22, Application US/08724394A
; Patent No. 5872237

; GENERAL INFORMATION:

```

: : : : :
: APPLICANT: Feder, John N.
: APPLICANT: Kronmal, Gregory S.
: APPLICANT: Lauer, Peter M.
: APPLICANT: Ruddy, David A.
: APPLICANT: Thomas, Winston
: APPLICANT: Tsuchinashi, Zenta
: APPLICANT: Wolff, Roger K.
: TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
: TITLE OF INVENTION: Sequences and Antibodies Thereto
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
: ADDRESSES: TOWNSEND and TOWNSEND and CREW LLP

```

STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-22

Query Match	3.6%	Score 199.2	DB 2	Length 246240
Best Local Similarity	61.7%	Pred. No. 2.9e-32		
Matches 439	Conservative 0	Mismatches 233	Indels 40	Gaps 6
QY	3199	TTTTAGATTCATTTATCTCTTTTATGTTTGTGTTTTTAAAGACAATCTCACCT	3258	
Db	3311		3352	
QY	3259	GTTACCCAGGCTGGAGTGCAGTGGTGCAGATCATAGATCTTCTGCACTCTTGAACCTCCGGG	3318	
Db	3251		3192	
QY	3319	CTCAAGCAATCCCTCCCTGGCTGCCCTCCCAAGTGTGGGA-TACAGTCAATGAGCCACTG	3377	
Db	3191	TTCAAGCCATCTCCCTGCCCTCAGCCTCCCAAGTAGCTGGACTACAGGCCGCCGCCACTA	3132	
QY	3378	CATCTGGC--CTAGGATCCATTTAGATTTAAATATGCATTTTAAATTTTAAATATATATG	3435	
Db	3131		3072	
QY	3436	GCTAATTTT-----ACCTTATGTAATGTGTATCT	3466	
Db	3071	GTCTCGATCTCTGACCTCGTGATCGCGCGCTCGGCCTCCCAAAGTCTGGATACA	3012	
QY	3467	GGTAATAATCTAGTTTGCTGCTAAAGTTTAAAGTGCTTTCCAATAAGCTTCATGTACG	3526	
Db	3011	GGGTGAGCCACCGCGCGGCTTATAAGACTTTTTAGATAAAACAAGATCTATATC	2952	
QY	3527	TGAGGGGAGACATTTAAAGTGAACACAGACAGCAGGTGTGGTGCACGCCCTGTAAATCC	3586	
Db	2951		2893	
QY	3587	CAGCACTCTGGAGGCTGAGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACCAGCCT	3646	
Db	2892	CAGCACTTTGGAGGCGGAGGCGGTAGATCACTTGAGTCAAGGATCGAGACCCAGCCT	2833	
QY	3647	GAGCAACATGCCAAACCTGTTTCT-----ATACAAAATTTAGCCGGCATGGTGC	3700	
Db	2832	GGCCAAACAGCGGAACCTGTCTCTACTAAAAATACAAAATTTAGCTGGCATGGTGC	2773	


```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 43950
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-735-934A-3

Query Match      3.5%; Score 193.8; DB 4; Length 43950;
Best Local Similarity 78.7%; Pred. No. 2.3e-31;
Matches 259; Conservative 0; Mismatches 62; Indels 8; Gaps 2;

QY 3536 ACATTTAAAGTGAACAGACAGACGCCAGGTGTGGCTCAGGCCTGTAATCCAGCACTCT 3595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14558 ACATCAAAATCTAGAAATCCAGCCAGGTGTGGCTCATGATGTAATCTCAGCACTTT 14617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3596 GGAAGCTGAGGTGGGTGGATGCCTTTAGCCCTGGAGTTCAAGACCCAGCCTGAGCAACAT 3655
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14618 GGGAGGCCAAGGTAGTGTGATCAGCTTGGAGCTTAGCAGTTTAAAGACCGCTGGGCCAAT 14677
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3656 GGCAAAACCCCTGTTTCTATAA-----CAAAATATAGCCGGGCATGTGGCCT 3708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14678 GGTGAACCCCTCTCTACAAAAAAATTAATAATTAGCCAGGCATGCGCATGCATCT 14737
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3709 GTGGTCCCCAGCTACTAGGG-GCCTGAGCAGGAGAAATCTTTGGAGCCCGAGGTCACAGG 3767
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14738 GTGGTTCAGACTACTTGGGACTCTGAGGAGGAGGATTCCTTGAGCCCGAGGTTGAGG 14797
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3768 CTGCACTGAGCAGTCTTGGCCCACTGCACTCCAGCCTGGGTGACAGGACCGACCTTGC 3827
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14798 CTGCACTGAGCCATGATTGTGCCACTGCATTCAGCCTGGGTGACAGAGCAAGAACTGT 14857
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3828 CTCAAAAAATAAAGAGAAAAATTAATAA 3856
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14858 CTCAAAAAAGAAAAAATAATTTCTAGA 14886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-09-791-211-10/c
; Sequence 10, Application US/09791211
; Patent No. 6448080
; GENERAL INFORMATION:
; APPLICANT: Donna T. Ward
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF WRN EXPRESSION
; FILE REFERENCE: RTS-0205
; CURRENT APPLICATION NUMBER: US/09/791.211
; CURRENT FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 90
; SEQ ID NO 10
; LENGTH: 9884
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 24962
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 64383
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 65468
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 65469
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 65470
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 65471
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 87130

```



```
Db 93229 CAGTCTCCCAAGTAGCTGAAATATACAGGTGTGCGCCACCATGCCAGGCTAAATTTTGTAT 93288
QY 3397 TTAGATTAAATATGCAATTTTAAATTTTAAATATATATGCGCTAAATTTTACCTTATGTA 3456
Db 93289 TTTTGTAGTGAGACAGGGGTTTCCACCATGGTGCCAGGCTGATCTGGAACCTCCTGACCTCA 93348
QY 3457 TGTGTATACTGGTATATAAATCTAGTTTGTCTGCTGCTAAAGTTTAAAGTGTCTTTCCCAATAGC 3516
Db 93349 AGTG-ATCCACCTGTGTGGCTTCCCAAGTGTCTGGAATTACAGGTGTGGCCACTGGC 93407
QY 3517 TT--CATGTAGTGAGGGGAGACATTTAAAGTGAACAGACAG-CCAGGTGTGGTGGCTC 3573
Db 93408 CTGGCCAGTTTATATTTTAAACATTTAAATATAAATGCTTCTTGTCCCAATCATGTGTGCTC 93467
QY 3574 ACGCTGTAAATCCCAAGCACTCTGGGAGGCTGAGGTGAGGTGGGTGATCGCTTGAGCCCTGGAGT 3633
Db 93458 ACACCCATTAATCCCAAGCACTTTGGGAGGCACAGCTTGGAGAACTCACCTTGAGGCCAGGAGT 93527
QY 3634 TCAAGACAGGCTGAGCAACATGCGCAAAACCCCTGTTCTATA-----ACAAAATTAGC 3687
Db 93528 TCGAGACCAAGCTGGGCAACATGTTGAATCCCTGCTCTACAAAAAATACAAAATTAGC 93587
QY 3688 CGGCATGGTGCATGTCCTGTGTGTCCTCCAGCTACTAGG-GGCTGAGGCAGGAGATCT 3746
Db 93588 TGGGCATGGTGTGTCATGCCCTGTAGTCCCAAGCTATTTGGGAGGCTGAGGTGGGAGATCA 93647
QY 3747 TTGGAGCCAGGAGGTCAAGGCTGACACTGACGAGTGTGGCGCCACTGCCTCCAGCCTG 3806
Db 93648 TTGAGCCCAAGATGTTGGGCTGCGAGTGAGCTGAGATCACCACTGACTTCCAGGCTG 93707
QY 3807 GGTGACAGGACCAAGCTTGCTGCTCAAAAAAATAAGAGAA 3846
Db 93708 GATGACAGCAAGGCCCGCTTTTCAAAAAAATAAGAGAA 93747
```

```
RESULT 14
US-09-754-250-3/c
; Sequence 3, Application US/09754250
; Patent No. 6376225
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHODIESTERASE
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: PHOSPHODIESTERASE PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001063
; CURRENT APPLICATION NUMBER: US/09/754,250
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 111282
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(111282)
; OTHER INFORMATION: n = A,T,C or G
US-09-754-250-3
```

```
Query Match 3.4%; Score 190.2; DB 4; Length 111282;
Best Local Similarity 74.8%; Pred. No. 1.7e-30;
Matches 267; Conservative 0; Mismatches 83; Indels 7; Gaps 2;

QY 3542 AAAGTGAACACAGACAGCAGGTGGTGGCTCAGCCCTGTAATCCCAGCACTCTGGGAGG 3601
Db 18630 AAAGTGAATGTTCAGGCCAGGCTGAGTGGCTCAGCCCTGTAATCCCAGCACTTTGGGAGG 18571
QY 3602 CTGAGGTGGGRTGATCGCTTGAGCCCTTGAGCCCTTCAAGACCAAGCCTGAGCAACATGGCAAA 3661
Db 18570 TCAAGGGCGGAGATCACTTGAAGGAGGAGTTCGAGACCAAGCCTGGCCATATATGGTCA 18511
QY 3662 ACCCTGTTCT-----ATACAAAAAATAGCCGGGCATGTGGCATGTGCTGTGGTCC 3715
```

```
Db 18510 ACCCGCTCTTACCAAAAGTACAAAAATTTAGCCGGCATGTGGCAGTGCCTGTGATCC 18451
QY 3716 CAGCTACTAGG-GGCTGAGCGAGGAGAACTTTTGGAGCCCGAGGAGGTCAAGGCTGCACT 3774
Db 18450 CAGTACTTTGGAGGCTGAGCGAGGAACTTACCTTTGAACCCAGGAGGTGGAGGTTCAGT 18391
QY 3775 GAGCAGTGTCTGGCCCACTGCACCTCCAGCCTGGGTGACAGGACGAGCCTTGGCCTCAAAA 3834
Db 18390 GAGACAAGATCGCACCACTGCACCTCCAGCCTGGCGACAGAGCAAGACTGTCTCAAAAA 18331
QY 3835 AAATAAGAAGAAAAATTTAAATAATATGGAACAACACTACAAAAAGCTGTTGCTCTAG 3891
Db 18330 AAAAAAAGAGAAAAAAGAAAAAACCACCAACAAAAAAGTGTTCAGAG 18274
```

```
RESULT 15
US-09-426-290-1/c
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Berglind Ran Olafsdottir
; APPLICANT: Jeffrey Gulcher
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; CURRENT FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21181)...(21403)
; NAME/KEY: CDS
; LOCATION: (95252)...(95430)
; NAME/KEY: CDS
; LOCATION: (101753)...(101996)
; NAME/KEY: CDS
; LOCATION: (110324)...(110439)
; NAME/KEY: CDS
; LOCATION: (124058)...(124278)
; NAME/KEY: CDS
; LOCATION: (127009)...(127130)
; NAME/KEY: CDS
; LOCATION: (128910)...(129139)
US-09-426-290-1
```

```
Query Match 3.4%; Score 188.2; DB 4; Length 168575;
Best Local Similarity 64.7%; Pred. No. 5.2e-30;
Matches 314; Conservative 0; Mismatches 163; Indels 8; Gaps 2;

QY 3402 TTAATAATATGCATTTTAAATTTTAAATAATATATGCTAAATTTTACCTTATGTAATGTT 3461
Db 88918 TTAATAAATTTTAAATGTTCTGAGGTGCGATAGTGACGATGTTGCACTATAAGGTGAAT 88859
QY 3462 ATACTGGTAAATAATCTAGTTTGTGCTTAAAGTTTAAAGTGTCTTCCAATAAGCTTCAT 3521
Db 88958 ATACTTAATGTTCATTGAACATATACACTTAAACACTGATTAGTTGTTGAATTTGTAATAT 88799
QY 3522 GTAGCTGAGGGGAGACATTTTAAAGTGAACACAGACGAGGCTGTGGTGGCTCAGCCCTGT 3581
Db 88798 GTATATTTTACCACAAATTTTAAATAATTAGTTGCCACTGGGAGTGGTGGCTCATGCCCTGT 88739
QY 3582 AATCCCAGCACTCTGGGAGGCTGAGGTGGGTGATGCTTGGAGCCCTGGAGTTCAGAGACC 3641
Db 88738 AATCCCAGCACTTTTGGGAGGCCGAGGTGGGTGGATCCCTCGATCAGATCAGGAGTTCGAGACC 88679
QY 3642 AGCCTGAGCAACATGGCAAAACCCCTGTTCT-----ATAACAAAAAATAGCCGGCAT 3694
Db 88678 AGCTGGCCAAACATAGTGAATCTCTCTACTATAAAAAAATACAAAAAATAGTGGGCTT 88619
```

QY	3695	GGTGCATGTGCTGTGTGCCAGCTACT-AGGGGGCTGAGCGAGGAATCTTTGGAGC	3753
Db	88618	GGTGTGGAGCCTGTAATCCAGCTACTCAGGAGCTGAGCGAGGAATCGCTTGAAC	88559
QY	3754	CCAGGAGTCAAGGCTGCACGTGAGCAGTGTGCGCCACTGCACTCCAGCCTGGGTGACA	3813
Db	88558	CCGGAGGTGGAGGCTGCAGTGCAGCCGAGATTTCACCACTGCACCTCCAGCCTGGCGACA	88499
QY	3814	GGACCAGACCTTGGCTCAAAAAAATAAGAGAAATTAATAATGGAACAACACTAC	3873
Db	88498	GAGTGAGGCTCCCTCTCACAAAAAATAAAAAAATTTAGTTGCCATAGTAAG	88439
QY	3874	AAAGA	3878
Db	88438	ACAGA	88434

Search completed: June 19, 2003, 11:14:04
Job time : 239.577 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 00:06:39 ; Search time 739.811 Seconds
(without alignments)
10937.110 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 5514
Sequence: 1 acagacgaatacatagtgcca.....tcaaactctgacctcagag 5514

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues
Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5512.8	100.0	5514	9	US-09-966-880A-9
2	4922.8	89.3	11204	9	US-09-966-880A-35
3	230.2	4.2	28818	10	US-09-764-877-2266
4	226.8	4.1	32204	9	US-09-764-872-517
5	226.8	4.1	32204	9	US-10-072-349-327
6	225.4	4.1	32204	10	US-09-764-855-327
7	225.4	4.1	28588	9	US-10-073-961-399
8	225.4	4.1	28588	10	US-09-764-887-399
9	221.6	4.0	10855	9	US-09-764-891-7965
10	221.2	4.0	32152	9	US-09-764-872-518
11	221.2	4.0	32152	9	US-10-072-349-328
12	221.2	4.0	32152	10	US-09-764-855-328
13	218.8	4.0	32134	9	US-09-764-891-6763
14	218.6	4.0	4421	9	US-10-216-373-6
15	215	3.9	19472	10	US-09-764-864-1698
16	214.4	3.9	17239	9	US-10-091-504-2326
17	214.4	3.9	17239	10	US-09-764-869-2326
18	213.6	3.9	32195	9	US-10-091-504-1605
19	213.6	3.9	32195	9	US-10-125-540-611

20	213.6	3.9	32195	9	US-10-125-540-617	Sequence 617, App
21	213.6	3.9	32195	10	US-09-764-870-611	Sequence 611, App
22	213.6	3.9	32195	10	US-09-764-870-617	Sequence 617, App
23	213.6	3.9	32195	10	US-09-764-869-1605	Sequence 1605, Ap
c 24	207.2	3.8	24132	9	US-10-074-095-661	Sequence 661, App
c 25	207.2	3.8	24132	10	US-09-764-860-661	Sequence 661, App
c 26	205.8	3.8	66686	10	US-09-736-960-86	Sequence 86, Appl
c 27	204.8	3.7	14769	9	US-09-764-891-7092	Sequence 7092, Ap
c 28	204.4	3.7	32195	9	US-10-091-504-1605	Sequence 1605, Ap
c 29	204.4	3.7	32195	9	US-10-125-540-611	Sequence 611, App
c 30	204.4	3.7	32195	9	US-10-125-540-617	Sequence 617, App
c 31	204.4	3.7	32195	10	US-09-764-870-611	Sequence 611, App
c 32	204.4	3.7	32195	10	US-09-764-870-617	Sequence 617, App
c 33	204.4	3.7	32195	10	US-09-764-869-1605	Sequence 1605, Ap
c 34	204.2	3.7	17996	10	US-09-764-877-2695	Sequence 2695, Ap
35	204	3.7	24132	9	US-10-074-095-661	Sequence 661, App
36	204	3.7	24132	10	US-09-764-860-661	Sequence 661, App
c 37	204	3.7	26657	10	US-09-810-673A-3	Sequence 3, Appl
c 38	202.8	3.7	32169	9	US-09-764-891-8605	Sequence 8605, Ap
c 39	202.8	3.7	52216	10	US-09-747-810-1	Sequence 1, Appl
40	201	3.6	7233	9	US-10-091-504-2123	Sequence 2123, Ap
41	201	3.6	7233	10	US-09-764-869-2123	Sequence 2123, Ap
42	200.6	3.6	23934	9	US-09-764-891-7210	Sequence 7210, Ap
c 43	200.6	3.6	23934	9	US-10-074-095-777	Sequence 777, App
c 44	200.6	3.6	23934	10	US-09-764-860-777	Sequence 777, App
45	200.6	3.6	23934	10	US-09-764-877-2536	Sequence 2536, Ap

ALIGNMENTS

RESULT 1
US-09-966-880A-9
; Sequence 9, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 5514
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: intron
; LOCATION: (1)...(1031)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1032)...(1118)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (1119)...(5514)
US-09-966-880A-9

Query Match 100.0%; Score 5512.8; DB 9; Length 5514;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5514; Conservative 0; Mismatches 0; Gaps 0;

Oy 1 ACACGACGAATACATGGTCCAAAGCTAGGCTATTGATTTTGAATAATCATCAAGGTATATAGTG 60
|||||

QY	2221	AACAAGTCTGAGCTATAGGTCCTTGTGAAGGTCCTAATGGAAATACATGTTGTTCAAGTAAA	2280		3301		3361	
QY	2221	AACAAGTCTGAGCTATAGGTCCTTGTGAAGGTCCTAATGGAAATACATGTTGTTCAAGTAAA	2280		QY	3301		3361
QY	2281	ATGGAAGCAAGAGTAAAATCAGCAGTTGAAATTCAGAGAAACACAGAAAGGAGAAAG	2340		QY	3361		3421
QY	2281	ATGGAAGCAAGAGTAAAATCAGCAGTTGAAATTCAGAGAAACACAGAAAGGAGAAAG	2340		QY	3361		3421
QY	2341	ATGAAATTCACAGGACAGAAAGGAAATATATATCATTTAAGGAGGACAGTATCTGTAGA	2400		QY	3421		3480
QY	2341	ATGAAATTCACAGGACAGAAAGGAAATATATATCATTTAAGGAGGACAGTATCTGTAGA	2400		QY	3421		3480
QY	2401	GCTCATATAGTATGGCAAAATGACTTGGTCAGAGTATATTTTAAACCCGCTGTTGTTCTGGT	2460		QY	3481		3540
QY	2401	GCTCATATAGTATGGCAAAATGACTTGGTCAGAGTATATTTTAAACCCGCTGTTGTTCTGGT	2460		QY	3481		3540
QY	2461	TTGCAGCGCTGGGGATGTCAGCTAGGTTCTGCGCTCAGGGAGCACAGCTGTCAGAGCAGC	2520		QY	3541		3600
QY	2461	TTGCAGCGCTGGGGATGTCAGCTAGGTTCTGCGCTCAGGGAGCACAGCTGTCAGAGCAGC	2520		QY	3541		3600
QY	2521	TGTCAGCGCTGCAAGCCTGAAACACTCCCTCGSTAAAGTCCCTTACTCAGGACAGAAAT	2580		QY	3601		3660
QY	2521	TGTCAGCGCTGCAAGCCTGAAACACTCCCTCGSTAAAGTCCCTTACTCAGGACAGAAAT	2580		QY	3601		3660
QY	2581	GACGAGAACAGGAGCTGGAACACAGGCCCTTAACGAGAGAGGAAATGATGATCAACA	2640		QY	3661		3720
QY	2581	GACGAGAACAGGAGCTGGAACACAGGCCCTTAACGAGAGAGGAAATGATGATCAACA	2640		QY	3661		3720
QY	2641	AAAGTTAACTAGCAGGTCAGGATCAGCAATTCATTTTCACTCTGACTGGTAACATGAC	2700		QY	3721		3780
QY	2641	AAAGTTAACTAGCAGGTCAGGATCAGCAATTCATTTTCACTCTGACTGGTAACATGAC	2700		QY	3721		3780
QY	2701	GAACAGTGTAGGCTTATTTGATTTTTCATGTAGATAGGACCCCAAAATCCACCCAAAGT	2760		QY	3781		3840
QY	2701	GAACAGTGTAGGCTTATTTGATTTTTCATGTAGATAGGACCCCAAAATCCACCCAAAGT	2760		QY	3781		3840
QY	2761	CCTTTATCTATGCCACATCCTTCTATCTATATCTTCCAGGACACTTTTCTTCTTATGA	2820		QY	3841		3900
QY	2761	CCTTTATCTATGCCACATCCTTCTATCTATATCTTCCAGGACACTTTTCTTCTTATGA	2820		QY	3841		3900
QY	2821	TAAGGCTCTCTCTCTCCACACACACACACACACACACACACACACACACACACAC	2880		QY	3901		3960
QY	2821	TAAGGCTCTCTCTCTCTCCACACACACACACACACACACACACACACACACACAC	2880		QY	3901		3960
QY	2881	CACAAACACACACCCCGCCCAACCAAGGTGCATGTAAAAGATGTAGATTCCCTGCGCTT	2940		QY	3961		4020
QY	2881	CACAAACACACACCCCGCCCAACCAAGGTGCATGTAAAAGATGTAGATTCCCTGCGCTT	2940		QY	3961		4020
QY	2941	CTCATCTACACGCCAGGAGGTAGTTAATAAAGAGGATTTATGGTAAGAGATGA	3000		QY	4021		4080
QY	2941	CTCATCTACACGCCAGGAGGTAGTTAATAAAGAGGATTTATGGTAAGAGATGA	3000		QY	4021		4080
QY	3001	TGCTTAATCTGTTTAACTGTCGCTCAAGAGAGAAATTTCTTCTCTGTACTTTATTA	3060		QY	4081		4140
QY	3001	TGCTTAATCTGTTTAACTGTCGCTCAAGAGAGAAATTTCTTCTCTGTACTTTATTA	3060		QY	4081		4140
QY	3061	AGCACCTATTATGTGTAGCTTATATACAAAGGTTATTTATGCTAATATAGTAAT	3120		QY	4141		4200
QY	3061	AGCACCTATTATGTGTAGCTTATATACAAAGGTTATTTATGCTAATATAGTAAT	3120		QY	4141		4200
QY	3121	AGTAATGKTGTTGGTACTATGTGTAATACCAATAAAATPAWTATCCTTTTAAATAAAG	3180		QY	4201		4260
QY	3121	AGTAATGKTGTTGGTACTATGTGTAATACCAATAAAATPAWTATCCTTTTAAATAAAG	3180		QY	4201		4260
QY	3181	CTAATTTATTTGATCTTTTTTATGATTTATTTTTTATGTTTTTATGTTTTT	3240		QY	4261		4320
QY	3181	CTAATTTATTTGATCTTTTTTATGATTTATTTTTTATGTTTTTATGTTTTT	3240		QY	4261		4320
QY	3241	AAAAGCAATCTCACCTGTTTACCCAGGCTGGAGTGCAGTGGTGAATCATAGCTTTCTG	3300		QY	4321		4380
QY	3241	AAAAGCAATCTCACCTGTTTACCCAGGCTGGAGTGCAGTGGTGAATCATAGCTTTCTG	3300		QY	4321		4380
QY	3301	CAGTCTTTGAACCTCTGGGCTCAAGCAATCCTCTGCTTGGCCTCCCAAGTGTGGGAT	3360		QY	4381		4440

```

Db 4381 TCTCACAAAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTTGTCTGTGGGAGGGAAG 4440
QY 4441 GACTGCAAGAGGGAAGAGCTCTGGTGGGTGAGGTGGTGAATTCAGGTCTGTATCCT 4500
Db 4441 GACTGCAAGAGGGAAGAGCTCTGGTGGGTGAGGTGGTGAATTCAGGTCTGTATCCT 4500
QY 4501 GACTGTGTACAGTTTGGGGGTGTTTACATCCAAAAATATTCGTPAGAATATGTCATCTTA 4560
Db 4501 GACTGTGTACAGTTTGGGGGTGTTTACATCCAAAAATATTCGTPAGAATATGTCATCTTA 4560
QY 4561 AATGGGTGGAGTTTACGTATGTAATAATATACCTCAATGTAAGAAAAAATAATGTGTAAG 4620
Db 4561 AATGGGTGGAGTTTACGTATGTAATAATATACCTCAATGTAAGAAAAAATAATGTGTAAG 4620
QY 4621 AAAAGTTTCAATCTCTGGCAGCAAAAGCTTATTCAAATTCCTGAGCCCTTTACTCGCA 4680
Db 4621 AAAAGTTTCAATCTCTGGCAGCAAAAGCTTATTCAAATTCCTGAGCCCTTTACTCGCA 4680
QY 4681 AATTCCTGCACTCTGCCCGGTACCAATAGGTGACAGCACTAGCTCCACAAATGGATA 4740
Db 4681 AATTCCTGCACTCTGCCCGGTACCAATAGGTGACAGCACTAGCTCCACAAATGGATA 4740
QY 4741 AATGCAATTTCTGGAAGAGACTAGGCAAAATCCAGGCATCACTTGCTTCATATCA 4800
Db 4741 AATGCAATTTCTGGAAGAGACTAGGCAAAATCCAGGCATCACTTGCTTCATATCA 4800
QY 4801 CCAGCTGTACAGCTTGTGTGCTGTCGAGCTGCAATGGGACTCTTGAATTTCTTAA 4860
Db 4801 CCAGCTGTACAGCTTGTGTGCTGTCGAGCTGCAATGGGACTCTTGAATTTCTTAA 4860
QY 4861 GGAACCTTGGGTTACCAGAGTATTTCCACAATGCTATTCAATAGTGCCTTATGATG 4920
Db 4861 GGAACCTTGGGTTACCAGAGTATTTCCACAATGCTATTCAATAGTGCCTTATGATG 4920
QY 4921 CAAGACACTGTGCTAGGAGCCAGAAACAAAGAGGAGGAGAAATCAGTCATTATGTTGGA 4980
Db 4921 CAAGACACTGTGCTAGGAGCCAGAAACAAAGAGGAGGAGAAATCAGTCATTATGTTGGA 4980
QY 4981 ACAACATACAGATATTTAGATCATTTTGAAGTATTTAAAGAGCAGAGTACAAAT 5040
Db 4981 ACAACATACAGATATTTAGATCATTTTGAAGTATTTAAAGAGCAGAGTACAAAT 5040
QY 5041 CACACATGCAATCAGTATAATCCAAATCATGTAATATGTCCTGTAGAAAGCTAGAGG 5100
Db 5041 CACACATGCAATCAGTATAATCCAAATCATGTAATATGTCCTGTAGAAAGCTAGAGG 5100
QY 5101 AATAACACAGAAATCTTAACAGTCAATGATAGACACTAAGTCTAATATTATTAAT 5160
Db 5101 AATAACACAGAAATCTTAACAGTCAATGATAGACACTAAGTCTAATATTATTAAT 5160
QY 5161 AGACACTATGATATTGAGATTTAAAAAATCTTTAATATTTTAAATTTAGAGCTCTCT 5220
Db 5161 AGACACTATGATATTGAGATTTAAAAAATCTTTAATATTTTAAATTTAGAGCTCTCT 5220
QY 5221 ATTTTTCATAGTATTCAAGTTTGACAATGATCAAGTATTAATCTTTCTTTTCTTTT 5280
Db 5221 ATTTTTCATAGTATTCAAGTTTGACAATGATCAAGTATTAATCTTTCTTTTCTTTT 5280
QY 5281 TTTTTTTTTTTTGGAGTGGAGTTTGGTCTTTGTTGCCATGCTGGAGTGAATGGCATG 5340
Db 5281 TTTTTTTTTTTTGGAGTGGAGTTTGGTCTTTGTTGCCATGCTGGAGTGAATGGCATG 5340
QY 5341 AYCATACTACTGCAACCTCCACTCTCTGGGTTCAAGCAAAAGTGTGCGCTCAGCCTCC 5400
Db 5341 AYCATACTACTGCAACCTCCACTCTCTGGGTTCAAGCAAAAGTGTGCGCTCAGCCTCC 5400
QY 5401 CGGTAGATGGGATTTACAGGCGCCACCACACACCTGGCTAATGTTGATTTTATGTA 5460
Db 5401 CGGTAGATGGGATTTACAGGCGCCACCACACACCTGGCTAATGTTGATTTTATGTA 5460
QY 5461 GAGATGGGTTTACCAATGTTGCCAGGCTGGTCTCAAACTCTGACCTCAGAG 5514
Db 5461 GAGATGGGTTTACCAATGTTGCCAGGCTGGTCTCAAACTCTGACCTCAGAG 5514

```

```

RESULT 2
US-09-966-880A-35
; Sequence 35, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 11204
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-35

```

```

Query Match      89.3%; Score 4922.8; DB 9; Length 11204;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 4921; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 591 AGTTTCAGAGAGACTGTGGGAATATGGGGAATATAGAGGCTATCTGAGGCTCTTCAACAC 650
Db 1 AGTTTCAGAGAGACTGTGGGAATATGGGGAATATAGAGGCTATCTGAGGCTCTTCAACAC 60
QY 651 AATAACCAAGAACTATTTAAATGCTCTTTAAAGGTATTTACATAAATATTAATATCTC 710
Db 61 AATAACCAAGAACTATTTAAATGCTCTTTAAAGGTATTTACATAAATATTAATATCTC 120
QY 711 ATTGTGCTTTTATTTGTGTATCATGATTATATAATGAAGTGTCTACTGTACTGCTCC 770
Db 121 ATTGTGCTTTTATTTGTGTATCATGATTATATAATGAAGTGTCTACTGTACTGCTCC 180
QY 771 TGATCTTTGTCTAGCTATGGAGCATGGACTGGGCTTTTAGAGCAGACGCCCAAGAAC 830
Db 181 TGATCTTTGTCTAGCTATGGAGCATGGACTGGGCTTTTAGAGCAGACGCCCAAGAAC 240
QY 831 TAAACATTAAGCAGAGCTGCCCTCAATGGTTTAACTGTGTGACTCTGCTCATGACAGC 890
Db 241 TAAACATTAAGCAGAGCTGCCCTCAATGGTTTAACTGTGTGACTCTGCTCATGACAGC 300
QY 891 CCCACCCACCATCTTCACTGGATCCAAATCAGGAGCAAGCGGTGGGTACCTGGTGG 950
Db 301 CCCACCCACCATCTTCACTGGATCCAAATCAGGAGCAAGCGGTGGGTACCTGGTGG 360
QY 951 GGTGTGTCTGTGAGGGGAGGAGGCCCAAGGCAAGCTCAAAATTTGAATGTGAAGGCC 1010
Db 361 GGTGTGTCTGTGAGGGGAGGAGGCCCAAGGCAAGCTCAAAATTTGAATGTGAAGGCC 420
QY 1011 AATGCACTGTGAGCTGAGCAGAGAACCATCATTAATTTGAAGTGTGAGATTTTCTGGCT 1070
Db 421 AATGCACTGTGAGCTGAGCAGAGAACCATCATTAATTTGAAGTGTGAGATTTTCTGGCT 480
QY 1071 GAGACTTGCAGGAGGAGCAAGAGACACTCTGACACCACTATGAGCAGTAAAGAGGCA 1130
Db 481 GAGACTTGCAGGAGGAGCAAGAGACACTCTGACACCACTATGAGCAGTAAAGAGGCA 540
QY 1131 TCTTCTGTTGGTGTATGTCAGCTGGCTTCTCTCAGAGCAAAATCTGAGTAATGAGACTGG 1190
Db 541 TCTTCTGTTGGTGTATGTCAGCTGGCTTCTCTCAGAGCAAAATCTGAGTAATGAGACTGG 600

```


Db	2761	GTGTTGGGATACAGTCAATGAGCCACTGCATCTGCGCTAGAGTCAATTTAGATTAAAAATAT	2820
Qy	3411	GCATTTTAAATTTAAATAATATATGGCTAAATTTTAACTTTATGTAAATGTCTATACCTGGTA	3470
Db	2821	GCATTTTAAATTTTAAATAATATATGGCTAAATTTTAACTTTATGTAAATGTCTATACCTGGTA	2880
Qy	3471	ATAAATCTAGTTTGGCTGCTTAAAGTTTAAAGTTTAAAGTCTTCCAAATAGCTTCAATGACGTGAG	3530
Db	2881	ATAAATCTAGTTTGGCTGCTTAAAGTTTAAAGTTTAAAGTCTTCCAAATAGCTTCAATGACGTGAG	2940
Qy	3531	GGGAGACATTTAAAGTGAACACAGACAGCCAGGTGTGGTGCTCACGCCCTGTATCCACGC	3590
Db	2941	GGGAGACATTTAAAGTGAACACAGACAGCCAGGTGTGGTGCTCACGCCCTGTATCCACGC	3000
Qy	3591	ACTCTGGGAGGCTGAGGTGGGTGGATCGCTTTGAGCCCTGAGGTTCAAGACCAGCCCTGAGC	3650
Db	3001	ACTCTGGGAGGCTGAGGTGGGTGGATCGCTTTGAGCCCTGAGGTTCAAGACCAGCCCTGAGC	3060
Qy	3651	AACATGGCAAAACCCCTGTTCTATACAAAAATTTAGCCGGGCATGGTGCGATGTGCTGT	3710
Db	3061	AACATGGCAAAACCCCTGTTCTATACAAAAATTTAGCCGGGCATGGTGCGATGTGCTGT	3120
Qy	3711	GGTCCAGCTACTAGGGGGCTGAGGCAGGAGAAATCTTTGAGCCCGAGGAGGTTCAAGGCTG	3770
Db	3121	GGTCCAGCTACTAGGGGGCTGAGGCAGGAGAAATCTTTGAGCCCGAGGAGGTTCAAGGCTG	3180
Qy	3771	CAGTGAGCAGTGTCTGGCCACTGCACTCCAGCTGGGTGACAGGACCAGACCTTGGCCTC	3830
Db	3181	CAGTGAGCAGTGTCTGGCCACTGCACTCCAGCTGGGTGACAGGACCAGACCTTGGCCTC	3240
Qy	3831	AAAAAATTAAGAGAAAAATTAATAATAATGGAACAACTACAAGAGCTTTGTCTCTA	3890
Db	3241	AAAAAATTAAGAGAAAAATTAATAATAATGGAACAACTACAAGAGCTTTGTCTCTA	3300
Qy	3891	GATGAGCTACTTAGTGTAGCTGATATTTTGGTATTAACTTTTAAAGTCAGGCTCTGTCA	3950
Db	3301	GATGAGCTACTTAGTGTAGCTGATATTTTGGTATTAACTTTTAAAGTCAGGCTCTGTCA	3360
Qy	3951	CTGCACATACATTAATAATATCAATCTCAATGTATATCCACAAGAGCTGTGACGT	4010
Db	3361	CTGCACATACATTAATAATATCAATCTCAATGTATATCCACAAGAGCTGTGACGT	3420
Qy	4011	GAATGTTTCATAGTACCTTTATTCACAAAACCCCAAGTAGAGACTATCCAAATATCCATC	4070
Db	3421	GAATGTTTCATAGTACCTTTATTCACAAAACCCCAAGTAGAGACTATCCAAATATCCATC	3480
Qy	4071	ACAAGTGAACAAATTAACAAAAATGTCTATATCCATGCAATGGAATACCCCTGTCAGT	4130
Db	3481	ACAAGTGAACAAATTAACAAAAATGTCTATATCCATGCAATGGAATACCCCTGTCAGT	3540
Qy	4131	ACAAAGGAAGAGCTTACTTGGGATGAATCCCAAGTTCATGCGCTAAATGAAGAGTCA	4190
Db	3541	ACAAAGGAAGAGCTTACTTGGGATGAATCCCAAGTTCATGCGCTAAATGAAGAGTCA	3600
Qy	4191	GACATGAAGAGGAGATATGTATGCCATACGAAATCTAGAAAATGAAAGTAACTTATA	4250
Db	3601	GACATGAAGAGGAGATATGTATGCCATACGAAATCTAGAAAATGAAAGTAACTTATA	3660
Qy	4251	GTTACAGAAAGCAATCAGGCGAGCATAGAGGCTCACACTGTAAATCCAGCACTTTGA	4310
Db	3661	GTTACAGAAAGCAATCAGGCGAGCATAGAGGCTCACACTGTAAATCCAGCACTTTGA	3720
Qy	4311	GAGGCCACGTGGGAAGATTGCTAGAACTCAGAGGTTCAAGACACAGCCCTGGGCAACAGT	4370
Db	3721	GAGGCCACGTGGGAAGATTGCTAGAACTCAGAGGTTCAAGACACAGCCCTGGGCAACAGT	3780
Qy	4371	GAAACTCCATCTCCACAAAATGGGAAAAAGAAAAGCAAAATCAGTGTGTCTGTGG	4430
Db	3781	GAAACTCCATCTCTCACAAAATGGGAAAAAGAAAAGCAAAATCAGTGTGTCTGTGG	3840
Qy	4431	GGAGGGAAGGACTGCAAGAGGGAAGCAAGCTCTGTGGGTGAGGCTGATTCAGGT	4490
Db	3841	GGAGGGAAGGACTGCAAGAGGGAAGCAAGCTCTGTGGGTGAGGCTGATTCAGGT	3900
Qy	4491	TCTGTATCTCTGACTGTGGTAGCAGTTTGGGTGCTTTTACATCCAAAAAATATTTCGTAGAAAT	4550
Db	3901	TCTGTATCTCTGACTGTGGTAGCAGTTTGGGTGCTTTTACATCCAAAAAATATTTCGTAGAAAT	3960
Qy	4551	ATGCATCTTTAAATGGGTGGAGTTTACTGTATGTAAATTTACTCTCAATGTAAAGAAAAAT	4610
Db	3961	ATGCATCTTTAAATGGGTGGAGTTTACTGTATGTAAATTTACTCTCAATGTAAAGAAAAAT	4020
Qy	4611	AATGTGTAAAGAAAAAGTTTCAATTCCTTCCAGCAAAACGTTATTTCAAATTCCTGAGCCCT	4670
Db	4021	AATGTGTAAAGAAAAAGTTTCAATTCCTTCCAGCAAAACGTTATTTCAAATTCCTGAGCCCT	4080
Qy	4671	TTACTTTCGCAAAATCTCTGCACTTCTGCCCCGTACCATTTAGGTGACAGCACTAGCTCCAC	4730
Db	4081	TTACTTTCGCAAAATCTCTGCACTTCTGCCCCGTACCATTTAGGTGACAGCACTAGCTCCAC	4140
Qy	4731	AAATTTGGATTAATGCAATTTCTGGAAGAACTAGGACAAAAATCCAGGCAATCACTTTGTGCT	4790
Db	4141	AAATTTGGATTAATGCAATTTCTGGAAGAACTAGGACAAAAATCCAGGCAATCACTTTGTGCT	4200
Qy	4791	TTTCATATCAACACACGCTGTACAGCTTGTGCTGTCTGCAAGTCAATTTGAGGACTCTTG	4850
Db	4201	TTTCATATCAACACACGCTGTACAGCTTGTGCTGTCTGCAAGTCAATTTGAGGACTCTTG	4260
Qy	4851	ATTTCTTTAAGGAAACTTTGGGTTACAGAGTATTTCCCAAAATGCTATTCAAAATTAGTGC	4910
Db	4261	ATTTCTTTAAGGAAACTTTGGGTTACAGAGTATTTCCCAAAATGCTATTCAAAATTAGTGC	4320
Qy	4911	TTATGATATGCAAGACACTGTGCTAGGAGCCAGAAAACAAAGAGGAGGAGAAATCAGTCA	4970
Db	4321	TTATGATATGCAAGACACTGTGCTAGGAGCCAGAAAACAAAGAGGAGGAGAAATCAGTCA	4380
Qy	4971	TTATGTTGGGAACAACATAGCAAGATATTTAGATCATTTTGTACTAGTTTAAAAAGCAGAG	5030
Db	4381	TTATGTTGGGAACAACATAGCAAGATATTTAGATCATTTTGTACTAGTTTAAAAAGCAGAG	4440
Qy	5031	AGTACAAAATCACACATGCAATCAGTATAATCCAAATCATGTAATATGTCCTGTAGAA	5090
Db	4441	AGTACAAAATCACACATGCAATCAGTATAATCCAAATCATGTAATATGTCCTGTAGAA	4500
Qy	5091	AGACTAGAGGAATTAACACAAGAAATCTTTAACAGTCAATTTGATACACTAAGTCTAAT	5150
Db	4501	AGACTAGAGGAATTAACACAAGAAATCTTTAACAGTCAATTTGATACACTAAGTCTAAT	4560
Qy	5151	TATTATTATTAGACACTATGATATTGAGATTTTAAAAAATCTTTAAATTTTAAAAATTTA	5210
Db	4561	TATTATTATTAGACACTATGATATTGAGATTTTAAAAAATCTTTAAATTTTAAAAATTTA	4620
Qy	5211	GAGCTCTCTATTTTCCATAGTATTCAAGTTTGCACAAATGATCAAGTATTACTCTTTCTT	5270
Db	4621	GAGCTCTCTATTTTCCATAGTATTCAAGTTTGCACAAATGATCAAGTATTACTCTTTCTT	4680
Qy	5271	TTTTTTTTTTTTTTTTTTTTTTTTTTTTCAGATGAGTTTTTGGTCTGTGGTCCCATGCTGAGTG	5330
Db	4681	TTTTTTTTTTTTTTTTTTTTTTTTTTTTCAGATGAGTTTTTGGTCTGTGGTCCCATGCTGAGTG	4740
Qy	5331	GAATGGCATGAYCATAGCTCACTGCAACCTCCAGCTCTCTGGTTTCAAGCAAAAGCTGTCG	5390
Db	4741	GAATGGCATGACCATAGCTCACTGCAACCTCCAGCTCTCTGGTTTCAAGCAAAAGCTGTCG	4800
Qy	5391	CTCAGCCCTCCGGGTAGATGGGATTACAGGCGCCCAACACACACTCGGCTAATTTTGT	5450
Db	4801	CTCAGCCCTCCGGGTAGATGGGATTACAGGCGCCCAACACACACTCGGCTAATTTTGT	4860
Qy	5451	ATTTTATTAGTAGAGTGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCTGACCTC	5510
Db	4861	ATTTTATTAGTAGAGTGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCTGACCTC	4920
Qy	5511	AGAG 5514	
Db	4921	AGAG 4924	


```
Db 18374 AGAATATTACAGTCACCGGTATATATGTGTGTGATATGTGTGCAAGACATATACCAGTTA 18315
Qy 5019 AAAAAGCAGCAGAGTAGTACAAATCACACATCAATCAATCAGTAAATCCAAATCATGTAATAT 5078
Db 18314 TCATTACATAGTGGGTATAGGCTAATATAAATGTTTATTCTTTGGGCTCTCTATCTTT 18255
Qy 5079 GTCCCTGTAGAAAGACTAGAGGAATAAACACAAGAATCTTAAACAGTCATGTGCAATAGAC 5138
Db 18254 -TCCAAGTTTCTTAAAGGCCCTGTATCTCTTTTACATTTGGGGTCAAGGCTGTTTTT 18196
Qy 5139 ACTAAGTCTAAATATTATTATTAGACATATGATATTTGAGATTTAAAAAACTTTAAATA 5198
Db 18195 ATTTTATTATTATTATTACAGCATGGCCACTATAATAATGTTATATATATGCGCTGAAT 18136
Qy 5199 TTTTAAAAATTTAGAGCTCTCTATTTTTCCATAGTATTCAAGTTTGCACA----- 5247
Db 18135 TTGCTTAAGAGAGATAAACCTGTTTTTCTTTTGCATCTCTTTTGTATTATTATTT 18076
Qy 5248 -----ATGATCAAGTATTACTCTTTCTTTT 5274
Db 18075 TGTGGGTATATAATAGTTGCTTGTATTATTATTATTATTATTATTATTATTATTATTAT 18016
Qy 5275 TTTTTTTTTTTTTTTTTTGGAGTGGAGTTTTGGTCTCTGTGCCATGCTGGAGTGAAT 5334
Db 18015 TTATTATTATTATTATTGAGATCGGAGTCTGCTCTGTTGCCAGGTTTGAGTGCAGT 17956
Qy 5335 GGCATGAYCATAGCTCACTGCAACCTCCACCTCCTGGGTTCAAGCAAAAGCTGTCGCCCTCA 5394
Db 17955 GCGCAATCTCAGCTCACTGCAACCTCCACCTCCTGGTTCAGCAATTTCTCCTGCCCTCA 17896
Qy 5395 GCTCCCGGTAGATGGGATTACAGCGCCGACCGCCACACACACACACACACACACACACAC 5454
Db 17895 GCCTCCGTAGCTGGGACTACAGCGACGACGACACACACACACACACACACACACACACAC 17836
Qy 5455 TTAGTAGAGATGGGTTTCAACATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCA 5511
Db 17835 TTAGTAGAGATGGGTTTCAACATGTTGGCCAGGCTGGTCTGCAACTCCTGACCTCA 17779

RESULT 4
US-09-764-872-517/c
; Sequence 517, Application US/09764872
; Publication No. US20030050231A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA125
; CURRENT APPLICATION NUMBER: US/09/764,872
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 957
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 517
; LENGTH: 32204
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-872-517

Query Match
Best Local Similarity 4.1%; Score 226.8; DB 9; Length 32204;
Matches 976; Conservative 0; Mismatches 967; Indels 59; Gaps 11;

Qy 3542 AAAGTGAACACAGACAGCCAGGTGGTGGCTCAGCCCTGTAATCCACAGCACTCTGGGAGG 3601
Db 18793 AAATTAATAAATAAATGCGAGTGGTGGCTCAGACCTGTAATCCACAGCACTCTGGGAGG 18734
Qy 3602 CTGAGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACCAGCCTGAGCAACATGGCAAA 3661
Db 18733 CCAAGACCGGTGGATCACTCAGTCAAGGAGTTCAGACACAGCCTGGCCAAACATGGTGA 18674
Qy 3662 ACCCTGTTTCT-ATACAAAAATATAGCCGGCATGGTGCATGTGCTGTGCTC 3715
Db 18673 ACCTCGTCTCTACTAAAAATACAAAAATTTAGCTGGGCGGTGTTGGGCATCTGTAATCC 18614
```

```
Qy 3716 CAGCTACT-AGGGGGCTGAGGCAGGAGAATCTTTTGGAGCCAGGAGGTCAAGGCTCGACT 3774
Db 18613 CAGCTACTCAGGAGGCTGAGGCAGGAGAATGGCTTGAGCCAGGAGGAGAGGTTGCGT 18554
Qy 3775 GAGCAGTGTGTGGCCACTGCACCTCCAGCTGGGTGACAGGACACACACCTTGCTCTCAAAA 3834
Db 18553 AAGCTCAGATGATGCCACTGCACCTCCAGCTCGAAGACAGAGTGAGACTCTGTCTCTCAAA 18494
Qy 3835 AAATAAGAAGAAAAATTAATAATTAATGAAACAACTACAAAAGAGCTGTTGCTCTAGATG 3894
Db 18493 ACAAAAAACAAACAAAAAATACAGATGGCCAGGCCAGGCATGCTGCTCACACC 18434
Qy 3895 AGCTACTTATTTAGGCTGA-----TATTTTGGTATTTAATTTTAAAGTCAGGGTCTGTC 3949
Db 18433 TGTATATACAGCACTATGGAAGGCTGAGGAGGAGGATTGCTTAAAGCCCTAGAGTTCCAGA 18374
Qy 3950 ACCTGCACCTACATATTATAAATATATCAATTTCTCAATGTATATCCACACAAGACTGGTAGC 4009
Db 18373 CCAGACTGGGCAATGTAGTGAGACCCCCCATCTCTACAGGAACAATAAAAATACANATA 18314
Qy 4010 TGAATGTTTCATAGTACCTTTTATTCACAAACCCCAAGTAGAGACTATCCAAATATPCCAT 4069
Db 18313 AAATGGCAGTTTCTCAGTTTGTCTGGGTTCTATGTGGTCACTGTCTCTTCAGAAATC 18254
Qy 4070 CAACAAGTGAACAAAAATAAACAAAAATGTCTATATCCATGCAATGGAATACCAACCCTGCAG 4129
Db 18253 ACTGATGGCCAGGTTCACTCTTCAAGACTGGGAAGCAGCAGCACAAGGATGGGATTTTCG 18194
Qy 4130 TACAAAGGAAGAAAGTACTTTGGGATGAATCCCAAGTCAATGACGCTAAATCAAAAGATC 4189
Db 18193 AGGCCAGAACTCCAGGTATTTTAAATTTTAAATTTTAAATTTATACAGGGGCCAAGC 18134
Qy 4190 AGACATGAAGGAGGAGATAATCTATGCCATAGCAAAATTTCTAGAAAAATGAAAGTAACCTAT 4249
Db 18133 ACGGTGGCTCACACCTGTAATTTCCGGTACTTTGGGAGAACATGCGCAAAACCCCTGCTCTA 18074
Qy 4250 AGTTACAGAAAGCAATCAGGCGAGCATAGAGGCTCACACTGTAATCCAGCACTTTG 4309
Db 18073 CTAATAATACAAAAAATCAGGCTGGCGCGGTAGCTCATGCTGTATATCCAGCACTTTG 18014
Qy 4310 AGAGGCCAGCT-GGGAAGATGCTAGAACTCAGGAGTTCAAGACCAGCCTGGGCAACACA 4368
Db 18013 GGAGGCCAAGGAGGTGGATCACCCTGAGATCAGGAGTCTGAGCCAGCCTGGCCAACTG 17954
Qy 4369 GTAAACTCCATTTCTCCACAAAAATGGGAAAAAGAAAGCAAAATCAGTGTGTGCTCTGT 4428
Db 17953 GCGAAACCCCATCTCTACTAAAAATACAAAAATTAGCCGGCATGGTTGTGTACACCTTT 17894
Qy 4429 ----GGGAGGGGAAGGACTCCAAGAGGGAAGAGCTCTGGTGGGTGAGGGTGGTGAAT 4484
Db 17893 AATCCAGCTACTAAGGAGGCTGAGGCAAGAGAATTTGTTGAACCTGGCAAGCGGTGTT 17834
Qy 4485 TCAGTGTCTGTATCTCTGACTGTGTAGCAGTTTGGGTGTTTACATCCAAAAATATTTCGT 4544
Db 17833 CGAGTAAGTCAAGATTGTGTGCCAATGCAATCCAGCCTGGGTGAGCAGCAAGACTCCGCT 17774
Qy 4545 AGAATTATGCACTTTAAATGGGTGGAGTTTACTGTATGTAATTTATACCTCAATGTAAGA 4604
Db 17773 CAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 17714
Qy 4605 AAAAATAATGTGAAGAAAAGTTTCAATTTCTCTTGGCAGCAACGTTTATTCAAAATTCCTG 4664
Db 17713 TGTGTTGGTGCACCTGTAGTCCAGTTTACTCAGGAGGCTGAGGTGGGAGGATCAGCTG 17654
Qy 4665 AGCCCTTTACTTCCCAAAATTTCTGCACTTCTGCCCCGTA-----CCATTAGG 4712
Db 17653 AGCCCGGAAGTTGAGGCTCGAGTGAGCTGTGATCGCACCACTTGCACACTACAGCCTGGGTG 17594
Qy 4713 TGACAGCACTAGCTCCCAAAATTTGGATTAATGCAATTTCTGGAAAGAACTAGGGACAAAAT 4772
Db 17593 ACAGAGCAAGACTCCGTTCTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATA 17534
```

QY 4773 CGAGCATCATTGTGCT-TTCATATCAACCCAGCTGTACAGCTTGTGTGTGTGCA 4831
DB 17533 TCACGTGAACCTTGAACCTTACTTCAAGAGCTCTACACAGCTGTGTGCC 17474
QY 4832 GCTCAATGGGACTCTTGAATTCCTTTAAGGAACTTGGTTACAGAGTATTTCCACA 4891
DB 17473 ACCCCCTGGCTAATTTTTTAAATTTTTTGTAGATGGGGTCAATGCTATGTTGCCCAGG 17414
QY 4892 ATGCTTTCAAAATTAGTCTTATGATATGCAAGACACTGTGTAGGAGCCAGAAAAACAA 4951
DB 17413 CTGCTCTCAAACTCTGGCCCAAGTGAGCTCTTGGCTTGGCTTCCCTAAGTGTGGGA 17354
QY 4952 GAGGAGGAGAAATCAGTCATTTATGTTGGGAACAACATAGCAAGATATTTAGATCATTTTGA 5011
DB 17353 TTATAGGTGGAGC-----CACTGTCATTCGGTTGTCATTTCT 17318
QY 5012 CTAGTTAAANAAGCAGCAGATACAAAATCACACATGCAATCAGTATATCCAAATCATG 5071
DB 17317 TTATTTAATACATCTTGATCTTTTCATATCAATATACATATCCCTCATATTAATGCTGTG 17258
QY 5072 TAAATATGTGCTGTAGAAAGACTAGAGGAATAAAACACAAGAAATCTTAACAGTCAATGTC 5131
DB 17257 CATACAACGATAGTGTGGGTAGTTATTTAGCTATCTTTTTTTCACGGAAA--TT 17200
QY 5132 ATTAGACACTAAGTCTAATATATATTATAGACACTATGATATTTGAGATTTTAAAAATC 5191
DB 17199 ATGATTTGTACAGTTTTTTGCTAATCAACAGAGTGTTCACAGTGAACATTTCTGTACAC 17140
QY 5192 TTTAATATTTTAAATTTAGAGCTCTTCTATTTTCCATAGTATTTCAAGTTTGACAATGA 5251
DB 17139 ACCCTTTATACATACATATGTTATTTATTTGAGAATAGAGAAGTGGGATTCCTGAGCTG 17080
QY 5252 TCAAGTATTAATCTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTAGATGGAGTTTGGTCT 5311
DB 17079 AAGATATGCAATTTGTTTTTTTGGGTTTTTTTTTTTTTTTTTTTTTTTGTAGATGGAGTTTGCCT 17020
QY 5312 TGTGCCCATGCTGGAGTGAATGCGATGAYCATAGCTACATGCAACCTCCACCTCCCTGG 5371
DB 17019 TGTGCCCAGGCTGGAGTAAAATGGCTGA-CTTGGCTCACTGAAACCTCTGCCCTCCCTA 16961
QY 5372 GTTCAAGCAAGCTGCTGCCCTCAGCTCCCGGTAGATGGGATTACAGCGCCGCCACC 5431
DB 16960 GTTCAAGCAATTTCTCGCTCAGCTCCCAAGTAGCTGGGATTACAAGCACCACCACC 16901
QY 5432 ACACTCGGCTAA--TGTTTGTATTTTAGTAGAGATGGGTTTACCACTATGTTGGCCAGGC 5489
DB 16900 AGCCCGAGCTAATTTTTTGTATTTTGGTAGAGCGGGTTTTACTAGTTGGCCAGGC 16841
QY 5490 TGGTCTCAAACTCTGACCTCA 5511
DB 16840 TGGTCTCAAACTCTCTGAACTCA 16819

RESULT 5

US-10-072-349-327
; Sequence 327, Application US/10072349
; Publication No. US20030054420A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P110C1
; CURRENT APPLICATION NUMBER: US/10/072,349
; CURRENT FILING DATE: 2002-02-11
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: Patentin Ver. 3.1
; SEQ ID NO 327
; LENGTH: 32204
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-072-349-327
Query Match 4.1%; Score 226.8; DB 9; Length 32204;

Best Local Similarity 48.8%; Pred. No. 5.8e-37;
Matches 976; Conservative 0; Mismatches 967; Indels 59; Gaps 11;
QY 3542 AAAGTGAACACACACAGCAGGTGTGTGCTCAGCGCTGAATCCACGACACTCTGGAGG 3601
DB 13412 AAATTAATAAAAAATGCCAGTGTGTGCTCACACTGTAAATCCACGACTCTGGAGG 13471
QY 3602 CTGAGTGTGGTGTGCTTGTAGCCCTTGAGCTTCAAGACAGCAGCTGAGCAACATGCAAA 3661
DB 13472 CCAAGCCGGTGTGATCACCTGAGGTCAAGAGTTCAGACAGCAGCTGCCCAACATGTGA 13531
QY 3662 ACCCTGTTTCT-----ATAACAAAAATTAGCCGGCATGTGTGGCATGTGCCCTGTGGTCC 3715
DB 13532 ACCTGCTCTCTACTAAAAATACAAAAATTAGTGGCGTGTGTGTGGCATCTGTAAATCC 13591
QY 3716 CAGCTACT-AGGGGGCTGAGCAGGAGAAATCTTTGGAGCCAGGAGGTCAAGGCTGCAC 3774
DB 13592 CAGCTACTCAGGAGGCTGAGGAGGAGAAATGGCTTGAGCCCGAGGAGGAGGTGCAGT 13651
QY 3775 GAGCAGTGTTCGGCCACTGCACTCCAGCTGGGTGACAGGACAGACCTTGCCTCAAAA 3834
DB 13652 AAGCTGAGATGATGCCACTGCACTCCAGCTCGAAGACAGAGTGAGACTCTGTCTCAAA 13711
QY 3835 AAATAGAAAGAAAAATTTAAAAATTAATGGAACAACACTACAAGAGCTGTTGTCTAGATG 3894
DB 13712 ACAAACCAACAAACAAAAATACAGATGGCCAGCCAGGCAATGGTGGCTCACACC 13771
QY 3895 AGCTACTTGTAGTGGCTGA- ----TATTTTGTATTTTAACTTTTAAAGTCAGGCTGTGC 3949
DB 13772 TGAATACAGCACTATGGAAGGCTGAGCAGGAGGATTCCTTAAGCCCTAGATGTTCAAGA 13831
QY 3950 AGCTGCACTACATATTAATAATATCAATTTCTCAATGTATATATCCACAAGACAGCTGATCG 4009
DB 13832 CCAGACTGGCAATGTAGTGAGCCCCCATCTCTACAGGAACAATAAAAACATAA 13891
QY 4010 TGAATGTTTCATAGTACCTTTTATACAAAACCCCAAGTAGAGACTATCAAAATATCCAT 4069
DB 13892 AAATGGCAGTTTCTCAGTTTGTGTGCTGTATGTGGTCACTGTCTCTTCAGAAATC 13951
QY 4070 CAACAGTCAACAAATAACAAAAATGTCTATATCCATGCAATGGAATACCACTTCGACG 4129
DB 13952 ACTGATGGCCAGGTTCACTCTTCAAGACTGGAAGCAGCAGCAAGGATGGGATTTTCG 14011
QY 4130 TACAAGGAAAGAGCTACTTGGGGATGAATCCCAAGTCAATGACGCTAAATGAAGAGTC 4189
DB 14012 AGGCCAGAACTCCAGTATTTTAAATTTTTTAAATTTTAAATTTACAGGGCCCAAGC 14071
QY 4190 AGACATGAAGGAGGAGATTAATGTATGCCATACGAAATTTAGAAAAATGAAGTAACTTAT 4249
DB 14072 ACGGTGGCTCACACCTGTAAATCCCGTACTTTTGGGAGAACATGGCAAAACCCCTGTCT 14131
QY 4250 AGTTACAGAAACAATCAGGCAGGCATAGAGGCTCACACCTGTATCCACGACTTTG 4309
DB 14132 CTAAAAATACAAAAATCAGGCTGGCGGGTAGCTCATGCTCTGTAAATCCACGACTTTG 14191
QY 4310 AGAGGCCACGT--GGGAAGATTGCTAGAACTCAGAGGTTCAAGACCAAGCTGGGCAACACA 4368
DB 14192 GGAGCCCAAGCAGGTGGATCACCTGAGATCAGGAGTCTGAGCCAGCCTGGCCACGCTG 14251
QY 4369 GTGAAACTCCATTTCCACAAAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTGTCTGCTGT 4428
DB 14252 GCGAAACCCCATCTCTACTAAAAATACAAAAATAGCCGGCATGGTGTGTGTACACCTTT 14311
QY 4429 ----GGGGAGGGAAGGAGCTGCAAGAGGGAAGAACTCTGGTGGGTGAGGTTGAT 4484
DB 14312 AATCCAGCTACTTAAGAGGCTGAGGCAAGAGAAATTCGTTGAACCTGGGAAGCGGTG 14371
QY 4485 TCAGGTTCTGTATCTCTGACTGTGGTAGCAGTTTGGGGTGTGTTTACATCCAAAAATATTCGT 4544
DB 14372 GCAGTAAGTCAAGATTGTGTCATTCGCAATCCAGCTGGGTGACAGCAAGACTCCGTCT 14431
QY 4545 AGAATTATGCATCTTAAATGGGTGGAGTTTACTGTATGTAATTAATCACTCAATGAAGA 4604


```

; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/241,785
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/244,617
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 60/225,268
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,368
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/251,856
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/251,868
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/229,344
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/234,997
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: 60/229,343
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,345
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,287
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,513
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/231,413
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/229,509
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/236,367
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/237,039
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,038
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/236,370
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/236,802
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,037
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,040
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/240,960
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/239,935
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/239,937
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/241,787
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,474
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,532
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,216
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,210
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/226,681
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,759
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/225,213
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/227,182
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,214
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/235,836
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/230,438
; PRIOR FILING DATE: 2000-09-06

```

```

; PRIOR APPLICATION NUMBER: 60/215,135
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/225,266
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,218
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,213
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,212
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,245
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,264
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08

```

Query Match 4.1%; Score 225.4; DB 9; Length 28588;
 Best Local Similarity 48.3%; Pred. No. 1.1e-36;
 Matches 1154; Conservative 1; Mismatches 1142; Indels 94; Gaps 15;

```

Oy 3213 TTTATGTTTTTATGTTTTTATGTTTTTAAAGACAAATCTCACCTCTTACCCAGGCTGG 3272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19521 TTTTATTATTATTATTATTTTATGATGGAGTTTCACTCTGTTCCCGGCGGG 19580

```


[illegible]


```

; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 328
; LENGTH: 32152
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-072-349-328

```

Query Match	4.08;	Score 221.2;	DB 9;	Length 32152;
Best Local Similarity	48.9%;	Pred. No. 8.8e-36;		
Matches 984;	Conservative 0;	Mismatches 968;	Indels 62;	Gaps 12;
Qy	3531	GGGAGACATTTAAAGTGAACACAGACGCCAGGTGTGTGGCTCACGCCTCTAATCCAGC	3590	
Db	8535	GGAAAAAATAATTAATAAANAATGCCAGATGTGTGGCTCACACTGTAATCCAGC	8594	
Qy	3591	ACTCTGGGAGGCTGAGTGGTGGATGCTTGGAGCCCTGGAGTTCAAAGACCGAGCCTGAGC	3650	
Db	8595	ACTCTGGGAGGCCAAGACCGGTGGATCACCTGAGGTCAGGAGTTCAAAGACCGAGCCTGGCC	8654	
Qy	3651	AACATGGCAAAACCCCTGTTTCT-----ATAACAAAAATTTAGCCGGCATGTGGCAATGT	3704	
Db	8655	AACATGGTGAACCTCGTCTCTACTAAAAATACAAAAATTTAGCTGGCGGTGGTTGGGC	8714	
Qy	3705	GCCTGTGGTCCCAAGCTACT-AGGSGGCTGAGGCAGGAGAAATCTTTGGAGCCACGAGAGTC	3763	
Db	8715	ATCTGTAAATCCAGCTACTCAGGAGCTGAGGCAGGAGAAATGGCTTGAGCCACGAGAGCA	8774	
Qy	3764	AAGCTGCACCTGAGCAGTGTCTGGCCACTGCACTCCAGCCTGGTGACAGGACCAAGCC	3823	
Db	8775	GAGGTTGCAGTAAGCTGAGATGATGCCACTGCACTCCAGCCTGGAAGACAGAGTGAGACT	8834	
Qy	3824	TTGCCTCAAAAAATAGAGAGAAAAATTAANAATAAATGANAACAACACTACAAGAGCTGT	3883	
Db	8835	CTGTCTCAAAAACAACAAAAACAACAACAAAAATACAGATGGCCAGGCCAGGCATGG	8894	
Qy	3884	TGCTCTAGATGAGCTACTTAGTTAGGCTGA-----TATTTTGGTATTTAACCTTTTAAAGT	3938	
Db	8895	TGGCTCACACCTGTAAATACAGCACTATGGAAGGCTGAGGCAGGAGGATTGCTTAAGCCCT	8954	
Qy	3939	CAGGCTGTCACTGCACCTGACATTAATTAANAATCAATTTCTCAATGTATATCCACAA	3998	
Db	8955	AGAGTTCAAGACACAGACTTGGGCAATGTAGTGAGACCCCGCATCTCTACAGAGAAACAATA	9014	
Qy	3999	AGACTGGTACGTGAATGTTTCATAGTACCTTTATTCACAAAAACCCCAAGTAGAGACTATC	4058	
Db	9015	AAATACAAATAAATGGCAGTTCTCAGTTTGTCTGTCTGGTGTCTATGTGGTCACCTGTCT	9074	
Qy	4059	CAAAATPCCATCAACAAGTGAACAAATAAACAANAATGTGTATATCATCCATGCAATGGAATA	4118	
Db	9075	TCATTCAAGAACTACTGATGGCCAGGTTTCACCTCTTCAAGACTTGAAGCAGCAGCACAGGA	9134	
Qy	4119	CCACCTGCAGTACAAGGAAGAAGCTACTTTGGGGATGAATCCAAAAGTCAATGACGCTAA	4178	
Db	9135	TGGGATTTTCGAGGCCAGAACTCCAGGTAATTTTAAATTTTTTTTAAATTTTAAATTTATAC	9194	
Qy	4179	ATGAAGAGTCCAGACATGAAGGAGGAGATAATGTATGCCATACGAAATTTCTAGAAATGA	4238	
Db	9195	AGGGCCACGACGGTGGCTCACACCTGTAAATCCGGTACTTTGGGAGAACATGGCAGAAA	9254	
Qy	4239	AAGTAACTTTATAGTTACAGAAAGCAATCAGGCGAGGCATAGAGGCTCACACCTGTAAATC	4298	
Db	9255	CCCTGTCTCTACTAAAAATACAAAAATCAGGCTGGCGCGGTAGCTCATGCTGTAAATC	9314	
Qy	4299	CCAGCACTTTGAGAGGCCAGTG- GGAAGATTTGCTAGAACTCAGGAGTTCAAGACCAAGCC	4357	
Db	9315	CCAGCACTTTGGGAGGCCAAGGCAAGGTGGATCACTGAGATCAGGAGTCTGAGGCCAGCC	9374	
Qy	4358	TGGCAACACAGTGAACCTCCATTTCTCCACAAAAATGGGAAAAAACAAGCAAAATCAGT	4417	
Db	9375	TGGCAACATGGGGAACCCCTCTCTACTAAANAATACAAANAATTAACCCGGGCAATGTTG	9434	
Qy	4418	GGTTGTCTCT- ---GGGAGGGGAAGGACTGCAAGAGGGGAAGAAAGCTCTGTTGGGTG	4473	

Db	9435	TTGACACCTTTAAATCCCAAGCTACTAAGGAGGCTGAGGCAAGAGAAATTCCTTTGAACCTGGG	9494
Qy	4474	AGGCTGGTGATTTCAGGTTCTGTATCCTGACTGTGGTAGCAGTTTGGGCTGTTCATACATCCA	4533
Db	9495	AAGCGGTGGTGCACAGCAAGTCAAGATTGTGTGCCATTGCAATCCAGCCTGGGTGCACAGCA	9554
Qy	4534	AAAATATTCGTAGAATATATGCATCTTAAATGGGTGGAGTTTACTGTATGTAATATPACC	4593
Db	9555	AGACTCCGCTCTCAAAAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	9614
Qy	4594	TCAATGTAAGAAAAAATAATGTGTAAAGAAAAAGTTTCAATTCCTTTGCCACCAACGTTAT	4653
Db	9615	AATTAGCTGGGTGTGTTGGTGTCTACCTGTAGTCCAGTTACTTCAGGAGGCTGAGGTGGG	9674
Qy	4654	TCAAAATTCCTGAGCCCTTTTACTTTCGCAAAATTCCTGCACTTCCTGCCCCG-----	4702
Db	9675	AGGATCACTGAGCCCGGAAGTTGAGGCTGCAGTGAGCTGATCGCACCATTTGCACATA	9734
Qy	4703	-TACCATTAGGTGACAGCACTAGCTCCACAAATTTGGATAAATGCAATTTCTGGAAAAAGACT	4761
Db	9735	CAGCTGGGTGCACAGCAAGACTCCGCTTAAAAAATAAATATTTTTTTTATAGGATAT	9794
Qy	4762	AGGACAAAAATCCAGGCATCACTTGTGCTT-TCATATCAACCAAGCTGTACAGCTTGTGT	4820
Db	9795	GTGATCACAGCTCACTGTAACCTTGAACCTTACTTCAAGCAGTCTTACACACAGCCTAC	9854
Qy	4821	TGCTGTCTGACGCTGCAATGGGACTCTTGTATTTCTTTAAGGAAACTTTGGTTTACCAGAG	4880
Db	9855	AGGTGTGTCACCCCTGGCTAAATTTTTTAAATTTTTTTGTAGAGATGGGGTCATGCTA	9914
Qy	4881	TATTTCCAAATGCTATTCAAAATTAGTGCTTATGATATGCAAGACACTGTGTAGGAGC	4940
Db	9915	TGTTGCCAGGCTGCTCTAAACTCTGTGCCCTCAAGTGAGCCTCTGCTTGGCCTCCCT	9974
Qy	4941	CAGAAAAAAGAGGAGGAGAAATCAGTCAATATATGTGGGNACAAACATAGCAAGATATTTA	5000
Db	9975	AAGTCTGGGATTAATAGGTGGAGC-----CACTGCATTCG	10010
Qy	5001	GATCATTTTGTACTAGTTAAAAAAGCAGCAGAGTACAAAAATCACACATGCAATCAGTATAA	5060
Db	10011	GTTCGATTTCTTTATTTAAATACATCTTCATCTTTTCATATACATATACATATCCTCAT	10070
Qy	5061	TCCAAATCATGTAATATGTGCTGTAGAAAGACTAGAGGAATAAAC-ACAAGAATCTTA	5119
Db	10071	ATTAAGTGTGATAGAACAGCATAGTGTGGGTAGTTATTTAGCTATATCTTTTTTTC	10130
Qy	5120	ACAGTCATATGTCATTTAGACACTAAGTCTAAATTTATTTATTTAGACACTATGATATTTGAG	5179
Db	10131	ACGGAATATGATTTGTTTACAGTTTTTTCAGTTTTTTCATTTACCAAGAGTGTTCAGACAGTGAACATT	10190
Qy	5180	ATTTAAAAAATCTTTAAATATTTAAAAATTTAGAGCTCTCTATTTTCCATAGTATTTCAA	5239
Db	10191	CTTTGACACCTCTTTATACATACATATGTGTATTTTGGAAATAGAGAAGT----GG	10246
Qy	5240	GTTTGGACAAATGATCAAGTATTTACTCTTTCTTTTTTTTTTTTTTTTTTTTTTTTGGAGATG	5299
Db	10247	GATTCGTGAGTGAAAGATATGCAATGTTGTTTTTTTGGTTTTTTTTTTTTTTTTTTTGGAGATG	10306
Qy	5300	GAGTTTTGGCTGTGTTGCCCATGTGGAGTGGGAATGGCATGATGATGATGATGATGATGATG	5359
Db	10307	GAGTTTTGCTCTCTGTTTGGCCAGGCTGGAGTAAAAATGGCGTGA-CTTGGCTCACTGGAACCC	10365
Qy	5360	TCCACCTCCTGGTTCAAGCAAGCTGTGCGCTCAGCCTCCCGGGTAGATGGGATTTACAG	5419
Db	10366	TCTGCCCTCCCTAGTTTCAAGCAATTCCTTCGCTCAGCCTCCCAAGTAGCTGGGATTTACAA	10425
Qy	5420	CGCCACACACACTCGGCTAA--TGTTTTGTTATTTTAGTAGAGATGGGTTTACCA	5477
Db	10426	GCACCCACACACGCCAGCTAAATTTTTTTGTTATTTTTTTTGGTAGAGACGGGGTTTTACTA	10485
Qy	5478	TGTTTGGCAGGCTGGTCTCAAACTCCTGACCTCA	5511

[illegible]


```
Db 3927 AATCCAGCTACTTGGAGGCTGAGGAGAGAAATCTTTGAACCCAGGAGCGAGGTT 3868
QY 3770 GCACTGAGAGTCTTTGCGCCACTGCACCTCCAGCTGGGTGACAGCAGACCACTTGCCT 3829
Db 3867 GCAGTGAGCGAGATCGCTCCACTGCACCTCCAGCCAGGTGACAGTGTGAGACTCCATCT 3808
QY 3830 C-AAAAAATAAGAAAGAAATTAATAAATTAATGGAACAACATCAAAAGAGCTGTGTCC 3888
Db 3807 CAAAAAATAAAGAAAGCCCGGTGCGGTGAGCTCAAGCTGTAAATCCAGCACCTTGGG 3748
QY 3889 TAGATGAGTACTTATAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGCTGTGT 3948
Db 3747 AGCCGAGGCGAGATCAGAGGTGAGGATCCAGACCATCTGCTGCTAACACGGTGA 3688
QY 3949 CACCTGCATACATTAATAAATAATCAATTTCTCAATGTATATCCACAAAGACGTGTAC 4008
Db 3687 ACCCGGTCTCTACTAAAA-----TACAAAAAATAGTCG 3651
QY 4009 GTGAATGTTTCATAGTACCTTTATTCACAAACCCCAAGTAGAGACTATCCAAATATCCA 4068
Db 3650 GCGGTGGTGGCGGAGCCTGTATTCACAGCTACTCGGGAGGTTGAGGCGAGAGATGGG 3591
QY 4069 TCAACAAGTGAACAATAAACAATAATGTGCTATATCCATGCAATGGAATACCCACCTGCCA 4128
Db 3590 TGAACCCGGGAGGAGAGCTTGCACTGAGCGAGATCGCGCACTGCACCTCCAGCTGGG 3531
QY 4129 GTACAAAGGAAGAGCTACTTGGGATCAATCCCAAGTCATGACGCTAAATGAAGAGT 4188
Db 3530 TGACAGCCAGACTCCATCTCAAGAANAANAATACTAAGAGGTTCTGCTGCCAG 3471
QY 4189 CAGACATGAAGGAGAGATATATGTCATACGAAATTTCTAGAAAATGAAGTAC--- 4245
Db 3470 ATGTGGTCTCTGGGTGGGTGCTGCTGCGGTGGTGGTCTCGAGACCAG 3411
QY 4246 --TTATATGTTACAAAGCAATCAGGCGAGGATAGAGGCTCACACCTTAATCCAGC 4303
Db 3410 GTGCGGATTCGTGTAGTCTGCGAGGCCAGTGTAGAGGCTCACACCTGTAATCCAGT 3351
QY 4304 ACTTTGAGAGGCC--ACGTGGGAAGATTGTAGAACTCAGAGTTCAAGACCAAGCTGGG 4362
Db 3350 ACTTTGAGAGGCTGAGTGGGAGGATCCTCTGAGCCAGAGTTCAAAACCAAGCTGGG 3291
QY 4363 AACACAGTGAATCCATTTCTCCAAATAATGGGAANAAGAAACAATCAGTGGTGG 4422
Db 3290 AATATAGTGAACGCTATCACTACAAAAGTTGTTTGTAGTTCACCGAGCATGTTGGCAC 3231
QY 4423 TCCTGTGGGAGGAGGAGTCAAGAGAGGAGAGTCTGGTGGGTGAGGCTGAGGCTGGT 4482
Db 3230 ATGCTGTAGTCCCACTACTCGAGAAGCTGAGATAGGTCACCTGAGGCTGGTAGGCTG 3171
QY 4483 ATTCTCTGTATCTGCTGAGTGTGGTAGCAGTTTGGGGTGTATACATCCAAAA 4537
Db 3170 AGCTCGAGTGAGCCAGATCGCACTACTGCACTCCAGCTCGGTGACAGAGTAAGACCC 3111
QY 4538 TATTCGTAGAATATGCACTTTAAATGGGTGAGTGTACTGTATGTAATTAATCACTCAA 4597
Db 3110 TGTCTCAACAAAAATAACAGGTTCTGTGCTGAGCTTTGCTCAGGAAGATAAGAC 3051
QY 4598 TGTAGAAAAAATAATGTGTAAGAAAAAGTTTCAATTTCTTCCAGCAAAAGTTATTCAA 4657
Db 3050 TTTCCCAACAGTCCGAAGCTAGGATATAGGGCTGAGGAGGACTCAAAACGTTGCAGTC 2991
QY 4658 ATTCCTGAGCCCTTTACTTCCAAATCTCTGCACCTCTGCCCTGACCATTAGTGACA 4717
Db 2990 AGAGAGACACCTATTTTGGTGAATATTTTGAATAGTTAAGTGGCTCCAGTACAAAC 2931
QY 4718 GCAGTAGTCCACAATTTGGATAAATGCATTTCTGGAAAAGACTAGGACAAATCCAGG 4777
Db 2930 TCCAGTTTCCAGAAAAGACTTAAAACTAGTATCTCGGAGATCCAAAGTCTTTAAAGGA 2871
QY 4778 CATCACTTGTGCTTTCATATCAACCGCTGACAGCTGTGTGTGCTGTGAGCTGCA 4837
Db 2870 TTTAGCTGATATCAGGCACTCTGGATGTAGCCTGGGGTTTGGCTGGATTCACCTCG 2811
```

```
QY 4838 ATGGGAGCTCTTGATTCTTTAAGAAACTTGGGTACCAGAGTATTTCCACAANTGCTA 4897
Db 2810 CTCAGTGATGATGCTGCTGTTTCCCTGGAACTTCCCTGTGGTTTGTAGTAAAGAGATGT 2751
QY 4898 TTCAATTAGTGTCTTATGATATCAAGACACTGTGCTAGGAGCCAGAAAACAAAGAGGAG 4957
Db 2750 GTGACCACTCCAGTGGGAAGTGAAGCTCTTAGTAGTTACATTTCTTAATTAAGAAAGATT 2691
QY 4958 GAGAAATCAGTCATTTATGTGGGAACAACATAGCAAGATATTTAGATCATTTTGCATGTT 5017
Db 2690 ATAAACAAGAAATCAACAAGCTGTCTACCTCCAGAGACACACCGTTCACCTTACAGTA 2631
QY 5018 AAAAAAGCAGAGTACAAAAATCACACATAGCAATAGTATATAATCCAAATCATGTAATA 5077
Db 2630 TCTAGCTTCCTCGACTCTTTTGTAAATCGTGGAGTATGAAGTGTGACCCAGAGAAGTAA 2571
QY 5078 TGTGCTGTAGAAGACTAGAGGAATAACACAAGAAATCTTAACAGTCAATTTGTCTATTAGA 5137
Db 2570 ATATGGAGCCGTTTATTTAGCAGTTTCTCTCCAGGATGACGCTGTGCTGTGTTTCT 2511
QY 5138 CACTAAGTCTAATTTATTTATTAGACACTATGATATTTGAGATTTTAAAAAATCTTTAAT 5197
Db 2510 GCCAGTTTTCAGTCTCTCTCTGCCCTTCATCTGTACTCTTGGCAGAGTCAGTAAC 2451
QY 5198 ATTTTAAATTTAGAGCTCTCTATTTTCCATAGTATTTCAAGTTTGACAATGATCAAGT 5257
Db 2450 CAACCACTAATAAATC---AAAGTTGTAGTAAAGACTTTATAACACATATGCCCTCTC 2394
QY 5258 ATTACTCTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGTCTGTTGC 5317
Db 2393 TGAGTTGTTGGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 2334
QY 5318 CCATGCTGAGTGGAAATGGCATGATGATGCTACTGCAACCTCCACCTCTCTGGGTTCAA 5377
Db 2333 CCAGGCTGAGGCAATGGCGGATCTTGGCTTACGGCAACCTCCGCTTCCGGGTTCAA 2274
QY 5378 GCAAAAGCTGTGCTCTCAGCTCCCGGTAGATGGGATTACAGGCGCCACCAACCACTC 5437
Db 2273 GTGATTCCTGCTTCAGCTCCCGAGTAGCTGGGATTACAGGCGCCACCAACCACTC 2214
QY 5438 GCTAATGTTTGTATTTTATAGATAGGTTTACACCTTTTACACCTGTTGGCAGGCTGCTCA 5497
Db 2213 GCTAAT--TTTGTATTTTATAGAGCGGATTTCACTATGTTGGTCAAGTTTGGTATCA 2155
QY 5498 AACTCCTGACCTCAG 5512
Db 2154 AACTCCGCGCTCAG 2140
```

RESULT 15

```
US-09-764-864-1698/c
; Sequence 1698, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR APPLICATION DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1698
; LENGTH: 19472
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-1698
```

```
Query Match 3.9%; Score 215; DB 10; Length 19472;
Best Local Similarity 61.9%; Pred. No. 1.3e-34;
Matches 394; Conservative 0; Mismatches 235; Indels 8; Gaps 3;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:09:48 ; Search time 226.945 seconds
(without alignments)
6475.074 Million cell updates/sec

Title: US-09-966-880A-9

Perfect score: 9659

Sequence: 1 acagacgaatacataggtcca.....tcaaaactcctgacctcagag 5514

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlp
-Q=Cyn2.1/USPTO_spool/US09966880/runat_14062003_175523_10297/app_query.fasta_1.9493
-DB=A_Geneseq_101002 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09966880 -RCGN_1_1_632 &runat_14062003_175523_10297 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	345.5	3.6	361	23	ABG68738 Human prostate spe
2	334	3.5	110	21	AAB43292 Human ORFX ORF3056
3	332	3.4	202	22	AAE11962 Human c-myb-relate
4	311	3.2	105	22	ABG11476 Novel human diagno
5	311	3.2	105	22	ABG12655 Novel human diagno
6	309	3.2	93	22	ABB10980 Human ovarian and/
7	309	3.2	93	22	AAM96679 Human reproductive
8	308	3.2	107	22	AAU31902 Novel human secret
9	307	3.2	110	23	ABP41552 Human ovarian anti
10	306	3.2	175	22	AAU31725 Novel human secret
11	306	3.2	241	22	ABB11464 Human neuronal thr
12	304	3.1	76	22	AAB75337 Human secreted pro
13	304	3.2	397	17	AAR95913 Neural thread prot
14	300.5	3.1	211	22	ABG12796 Novel human diagno
15	298	3.1	112	21	AAG03789 Human secreted pro
16	294.5	3.0	78	22	ABB11677 Human secreted pro
17	294	3.0	73	22	AAO08565 Human polypeptide
18	294	3.1	304	22	AAU30391 Novel human secret
19	294	3.0	341	22	AAO12925 Human polypeptide
20	293	3.0	119	22	AAM83429 Human immune/haema
21	290	3.0	75	21	AAB42504 Human ORFX ORF2268
22	289	3.0	117	23	AAU11247 Human neuronal thr
23	288.5	3.0	136	23	AAE14727 Human Alzheimer's
24	288	3.0	84	22	AAM85552 Human immune/haema
25	288	3.0	108	21	AAG00352 Human secreted pro
26	287.5	3.0	110	21	AAB43292 Human ORFX ORF3056
27	287	3.0	126	23	ABG65578 Human breast speci
28	287	3.0	132	22	ABG08282 Novel human diagno
29	287	3.0	225	22	AAU31713 Novel human secret
30	286	3.0	89	21	AAB43207 Human ORFX ORF2971
31	286	3.0	94	23	ABJ03693 Human ovary specif
32	286	3.0	132	22	AAU69854 Human prostate cdn
33	286	3.0	132	22	AAU01209 Human prostate-spe
34	286	3.0	132	23	ABB95314 Human p776p splice
35	286	3.0	135	22	AAU69829 Human prostate cdn
36	286	3.0	135	22	AAU01284 P789p splice varia
37	286	3.0	135	23	ABB95389 Novel human secret
38	285.5	3.0	153	22	AAU31783 Human zinc finger
39	285	3.0	136	22	AAG78828 Novel human secret
40	284	2.9	81	22	AAU32132 Novel human secret
41	284	2.9	209	22	AAO12667 Human polypeptide
42	283	2.9	168	22	ABG07628 Novel human diagno
43	282.5	2.9	84	23	ABJ03718 Human ovary specif
44	282.5	2.9	97	23	AAE14726 Human Alzheimer's
45	282	2.9	160	22	AAG78758 Human ribosomal pr

ALIGNMENTS

RESULT 1

ABG68738

ID ABG68738 standard; Protein; 361 AA.

AC ABG68738;

DT 07-OCT-2002 (first entry)

XX Human prostate specific protein DEX0293_107.

DE Human; prostate specific nucleic acid; PSNA; prostate cancer; PSP;
KW prostate specific protein; cytostatic; non-cancerous prostate disease;
KW gene therapy; cancer; immunostimulant; vaccine.

OS Homo sapiens.

XX

PN WO200255735-A2.

XX 18-JUL-2002.

PD

XX PF 27-NOV-2001; 2001WO-US44363.
XX PR
XX PR 27-NOV-2000; 2000US-253176P.
XX PA (DIAD-) DIADEXUS INC.
XX PI Salceda S, Macina RA, Recipon H, Cafferkey R, Ali S, Sun Y;
XX PI Liu C, Chen S;
XX XX WPI: 2002-557831/59.
XX DR N-PSDB; ABK97631.
XX PT New prostate specific genes, useful for treating or diagnosing cancer,
XX PT or useful as vaccines for treating cancer, particularly prostate
XX PT cancer, in a patient
XX PS Claim 11; Page 207-208; 212pp; English.
XX CC The invention relates to a new isolated prostate-specific nucleic acid
XX CC (PSNA) molecule comprising the cDNA sequences appearing as ABK97574-
XX CC ABK97642 which encode prostate specific proteins appearing as
XX CC ABG68701-ABG68746, or a sequence hybridising to a PSNA or which has 60%
XX CC sequence homology with a PSNA. Also included are a method of determining
XX CC the presence of a PSNA in a sample, a vector comprising the PSNA,
XX CC a host cell comprising the vector, producing the polypeptide encoded by
XX CC the PSNA, a method of determining the presence of a PSP in a sample,
XX CC diagnosing and monitoring the presence and metastases of prostate cancer
XX CC in a patient, a kit for detecting a risk of cancer or presence of cancer
XX CC in a patient (the kit comprising a means for determining the presence of
XX CC the PSNA or PSP in a sample of a patient) and a vaccine comprising the
XX CC polypeptide or the nucleic acid encoding the polypeptide. The PSNA, PSP
XX CC and anti-PSP antibody are useful for diagnosing and treating cancer in a
XX CC patient (e.g. by gene therapy). The nucleic acid molecule and polypeptide
XX CC are also useful as vaccines for treating cancer, particularly prostate
XX CC cancer and non-cancerous prostate diseases. The present sequence is
XX CC a PSP of the invention.
XX SQ Sequence 361 AA;
Alignment Scores:
Pred. No.: 1-54e-30 Length: 361
Score: 345.50 Matches: 64
Percent Similarity: 71.8% Conservative: 5
Best Local Similarity: 66.67% Mismatches: 24
Query Match: 3.58% Indels: 3
DB: 23 Gaps: 1
US-09-966-880A-9 (1-5514) x ABG68738 (1-361)
QY 5235 TTCAGTGTGACATGATCAAGTACTATCTC-----TTCTTTTTTTTTTTTTTTT 5285
Db 266 PheLysPheAlaAspSerHisIleGlyLeuAlaPheHisPheAlaPhePhePhePhe 285
QY 5286 TTTTITTTGAGATGAGTTTGGTCTTCTTCCCATGCTGAGTGAATGTCATGAYCAT 5345
Db 286 PhePhePheAlaValAlaSerHisProIleAlaGlnAlaGlyValGlnTrpArgAspLeu 305
QY 5346 AGCTCACTGCACCTCCACCTCCTGGTTCACGAAAGCTGTGGCTCAGCTCCCGGT 5405
Db 306 GlySerLeuGlnProProProGlyPheLysGlnPheLeuCysLeuSerLeuProGly 325
QY 5406 AGATGGGATTACAGCGCCACCCACCATCTGGCTAATGTTGTTATTTAGTAGAT 5465
Db 326 SerTrpAspTyrArgAlaProProArgGlnAlaAsnPheCysIlePheSerArgAsp 345
QY 5466 GGGGTTTCCACCATGTGCGCAGGTGGTCTCAAACTCTGACCTCAGA 5513
Db 346 GlyValSerProCysTrpThrGlyTrpSerGlnThrProAspLeuArg 361

RESULT 2
AAB43292
ID AAB43292 standard; Protein; 110 AA.

XX AAB43292;
XX AC
XX DT 08-FEB-2001 (first entry)
XX DE Human ORFX ORF3056 polypeptide sequence SEQ ID NO:6112.
XX KW Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflamatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflamatory disease; coagulation;
KW thrombosis; contraceptive.
XX OS Homo sapiens.
XX PN WO2000058473-A2.
XX XX 05-OCT-2000.
XX 31-MAR-2000; 2000WO-US08621.
XX 31-MAR-1999; 99US-0127607.
XX PR 02-APR-1999; 99US-0127636.
XX PR 05-APR-1999; 99US-0127728.
XX PR 30-MAR-2000; 2000US-0540763.
XX (CURA-) CURAGEN CORP.
XX Shlnkets RA, Leach M;
DR WPI: 2000-602362/57.
DR N-PSDB; AAC77501.
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX Claim 11; Page 5297-5298; 5507pp; English.
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnary;
XX antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antinflamatory; antibacterial; antiviral; antifungal; antirheumatic;
XX antithyroid; and antianaemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX erythematosus, hypothyroidism, cholesterol ester storage, systemic lupus
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antinflamatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX Sequence 110 AA;
Alignment Scores:

FT	Misc-difference	20	/label= Unknown	/note= "Encoded by AGN of the inverse complementary strand of the sequence shown in AAD19493."
FT				
FT				
FT				

Pred. No.: 4,32e-29 Length: 202
Score: 332.00 Matches: 73
Percent Similarity: 75.70% Conservative: 8
Best Local Similarity: 68.22% Mismatches: 21
Query Match: 3,45% Indels: 6
DB: 22 Gaps: 3

US-09-966-880A-9 (1-5514) x AAEL1962 (1-202)

QY 3866 TTTCATTATTTTAAATTTCTTCTATTTTGGAGCAAGGCTGGTCTCT---GTC 3810
DB 78 PheTyrPheLeuPhe---PheLeuLeuPhePheTrpGlyArgSerValProLeuVal 96
QY 3809 ACCAGGTGGAGTGCAGTGGCGCAAGCACTGCTCAGTCAGCCTTGACCTCTCTGGGCTC 3750
DB 97 AlaGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnProSerProGlyPhe 116
QY 3749 CAAAGATTCTCTCGCTCAGCCC-CCTAGTAGCTGGACACAGGACATGCCACCATGC 3691
DB 117 Lys***PheSerCysLeuSerLeuProSerSerTrpAspTyrArgHisAlaProCys 136
QY 3690 CCGGCTAATTTTGT---ATAGAAACAGGTTTTCGCATGTTGCTCAGGCTGGTCT 3637
DB 137 ProAlaAsn-PheValPheLeuValGluThrGlyPheHisHisIleGlyGlnAlaGlyLe 156
QY 3636 TGAATCCAGGGCTCAAGCATCCACCCACCTCAGCTCCAGTCCAGTGGATTACAGG 3577
DB 156 uGluLeuGlnThrSerGlyAspProProThrSerAlaSerGlnThrAlaGlyIleThrG 176
QY 3576 CGTGAGCCACACACCTGG 3558
DB 176 yValSerHisArgAlaTrp 182

RESULT 4

ID ABG11476 standard; Protein; 105 AA.

XX AC ABG11476;

DT 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #11467.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS75663.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

PS Claim 20; SEQ ID No 41835; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 105 AA;

Alignment Scores:

Pred. No.: 8,84e-27 Length: 105
Score: 311.00 Matches: 63
Percent Similarity: 68.81% Conservative: 12
Best Local Similarity: 57.80% Mismatches: 26
Query Match: 3,22% Indels: 8
DB: 22 Gaps: 3

US-09-966-880A-9 (1-5514) x ABG11476 (1-105)

QY 5199 TTTTAAATTTAGAGCTCTTCTATTTCCATAGTATTCAAGTTTCACAATGATCAAGTA 5258
DB 1 PheLeuSerLeuProThrPheLeuPhe---ValIlePheSerGlyGluGluLeuLeu 19
QY 5259 TTACTCTTTCTT-----TTTTTTTTTTTTTTTTTTTTTTTTCAGATGGAGTTT 5306
DB 20 ValLeuAlaLeuValPheLeuSerLeuPhePhePhePheLeuArgTrpSerPhe 39
QY 5307 GGTCTTGTGTCCTGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGT 5366
DB 40 AlaValAlaValGlnAlaValValGlnTrpHisAsnLeuSerSerLeu-----Pro 56
QY 5367 CCGTGGTTCAAGCAAGCTGTGCGCTCAGCTCCCGGCTAGATGGATTACAGGCGCCCA 5426
DB 57 PheGlyPheLeuGlnPheSerCysLeuSerLeuProSerSerTrpAspTyrArgCysPro 76
QY 5427 CCACCACACTCGGCTAATTTGTATTTTATAGTAGATGGGTTTTCACCATGTTGGCCA 5486
DB 77 ProProArgProAlaAsnPheCystIlePheThrArgAspGlyValSerProCysCysPro 96
QY 5487 GCGTGGTCTCAACTCTCGACTCAGA 5513

DB 97 GlyTrpSerArgThrSerAspLeuArg 105

RESULT 5

ABG12655

ID ABG12655 standard; Protein; 105 AA.

XX AC ABG12655;

DT 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #12646.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

QY	5393	CAGCCTCCGGGTAGATGGGATTACAGCGCGCCACACACACTCGGCTAATGTTGTAT	5452	PR	08-SEP-2000;	2000US-0231242.
Db	43	userLeuLeuSerSerTyrArgHisThrProHisAlaAlaAsnPheCysI1 63		PR	08-SEP-2000;	2000US-0231243.
QY	5453	TTTTAGTAGACATGGGTTTCACCATGTTGCCAGGCTGGTCAAACTCCTGACCTC	5510	PR	08-SEP-2000;	2000US-0231244.
Db	63	epheserArgaspGlyValSerProCysTrpSerGlyTrpSerArgThrProAspLeu 82		PR	08-SEP-2000;	2000US-0231413.
RESULT 7				PR	08-SEP-2000;	2000US-0231414.
AA96679				PR	08-SEP-2000;	2000US-0232080.
ID	AA966679	standard; Protein; 93 AA.		PR	12-SEP-2000;	2000US-0231968.
AC	AA966679;			PR	14-SEP-2000;	2000US-0232397.
XX				PR	14-SEP-2000;	2000US-0232397.
DT	21-NOV-2001	(first entry)		PR	14-SEP-2000;	2000US-0232398.
XX				PR	14-SEP-2000;	2000US-0232399.
DE		Human reproductive system related antigen SEQ ID NO: 5337.		PR	14-SEP-2000;	2000US-0232400.
XX				PR	14-SEP-2000;	2000US-0232401.
KW		Human; reproductive system related antigen; reproductive system disorder;		PR	14-SEP-2000;	2000US-0233063.
KW		cancer; gene therapy.		PR	14-SEP-2000;	2000US-0233063.
XX				PR	14-SEP-2000;	2000US-0233064.
OS		Homo sapiens.		PR	14-SEP-2000;	2000US-0233065.
XX				PR	14-SEP-2000;	2000US-0233065.
PN	WO200155320-A2.			PR	21-SEP-2000;	2000US-0234223.
XX				PR	21-SEP-2000;	2000US-0234274.
PD	02-AUG-2001.			PR	25-SEP-2000;	2000US-0234997.
XX				PR	25-SEP-2000;	2000US-0234997.
PF	17-JAN-2001;	2001WO-US01339.		PR	26-SEP-2000;	2000US-0235484.
XX				PR	26-SEP-2000;	2000US-0235484.
PR	31-JAN-2000;	2000US-0179065.		PR	27-SEP-2000;	2000US-0235834.
PR	04-FEB-2000;	2000US-0180628.		PR	27-SEP-2000;	2000US-0235836.
PR	24-FEB-2000;	2000US-0184664.		PR	29-SEP-2000;	2000US-0236327.
PR	02-MAR-2000;	2000US-0186350.		PR	29-SEP-2000;	2000US-0236367.
PR	16-MAR-2000;	2000US-0189874.		PR	29-SEP-2000;	2000US-0236368.
PR	17-MAR-2000;	2000US-0190876.		PR	29-SEP-2000;	2000US-0236369.
PR	18-APR-2000;	2000US-0198123.		PR	29-SEP-2000;	2000US-0236370.
PR	19-MAY-2000;	2000US-0205515.		PR	02-OCT-2000;	2000US-0236802.
PR	07-JUN-2000;	2000US-0209467.		PR	02-OCT-2000;	2000US-0237037.
PR	28-JUN-2000;	2000US-0214886.		PR	02-OCT-2000;	2000US-0237038.
PR	30-JUN-2000;	2000US-0215135.		PR	02-OCT-2000;	2000US-0237039.
PR	07-JUL-2000;	2000US-0216647.		PR	13-OCT-2000;	2000US-0237040.
PR	07-JUL-2000;	2000US-0216880.		PR	13-OCT-2000;	2000US-0239935.
PR	11-JUL-2000;	2000US-0217487.		PR	13-OCT-2000;	2000US-0239937.
PR	11-JUL-2000;	2000US-0217496.		PR	20-OCT-2000;	2000US-0240960.
PR	14-JUL-2000;	2000US-0218290.		PR	20-OCT-2000;	2000US-0241221.
PR	26-JUL-2000;	2000US-0220963.		PR	20-OCT-2000;	2000US-0241785.
PR	14-AUG-2000;	2000US-0220964.		PR	20-OCT-2000;	2000US-0241786.
PR	14-AUG-2000;	2000US-0224518.		PR	20-OCT-2000;	2000US-0241787.
PR	14-AUG-2000;	2000US-0224519.		PR	20-OCT-2000;	2000US-0241808.
PR	14-AUG-2000;	2000US-0225213.		PR	20-OCT-2000;	2000US-0241809.
PR	14-AUG-2000;	2000US-0225214.		PR	01-NOV-2000;	2000US-0241826.
PR	14-AUG-2000;	2000US-0225266.		PR	08-NOV-2000;	2000US-0244617.
PR	14-AUG-2000;	2000US-0225267.		PR	08-NOV-2000;	2000US-0246474.
PR	14-AUG-2000;	2000US-0225268.		PR	08-NOV-2000;	2000US-0246475.
PR	14-AUG-2000;	2000US-0225270.		PR	08-NOV-2000;	2000US-0246476.
PR	14-AUG-2000;	2000US-0225271.		PR	08-NOV-2000;	2000US-0246477.
PR	14-AUG-2000;	2000US-0225447.		PR	08-NOV-2000;	2000US-0246478.
PR	14-AUG-2000;	2000US-0225757.		PR	08-NOV-2000;	2000US-0246523.
PR	14-AUG-2000;	2000US-0225758.		PR	08-NOV-2000;	2000US-0246524.
PR	18-AUG-2000;	2000US-0225759.		PR	08-NOV-2000;	2000US-0246525.
PR	22-AUG-2000;	2000US-0226279.		PR	08-NOV-2000;	2000US-0246526.
PR	22-AUG-2000;	2000US-0226681.		PR	08-NOV-2000;	2000US-0246527.
PR	22-AUG-2000;	2000US-0226688.		PR	08-NOV-2000;	2000US-0246528.
PR	23-AUG-2000;	2000US-0227182.		PR	08-NOV-2000;	2000US-0246532.
PR	30-AUG-2000;	2000US-0227009.		PR	08-NOV-2000;	2000US-0246609.
PR	01-SEP-2000;	2000US-0228924.		PR	08-NOV-2000;	2000US-0246610.
PR	01-SEP-2000;	2000US-0229287.		PR	08-NOV-2000;	2000US-0246611.
PR	01-SEP-2000;	2000US-0229343.		PR	08-NOV-2000;	2000US-0246613.
PR	01-SEP-2000;	2000US-0229344.		PR	17-NOV-2000;	2000US-0249207.
PR	01-SEP-2000;	2000US-0229345.		PR	17-NOV-2000;	2000US-0249208.
PR	05-SEP-2000;	2000US-0229509.		PR	17-NOV-2000;	2000US-0249210.
PR	05-SEP-2000;	2000US-0229513.		PR	17-NOV-2000;	2000US-0249211.
PR	06-SEP-2000;	2000US-0230437.		PR	17-NOV-2000;	2000US-0249212.
PR	06-SEP-2000;	2000US-0230438.		PR	17-NOV-2000;	2000US-0249213.

PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;
 XX

DR WPI; 2001-465570/50.
 DR N-PSDB; AAL02649.

XX Isolated nucleic acid molecule encoding a reproductive system antigen -
 PT is used in preventing, treating or ameliorating a medical condition -
 XX

PS Claim 11; SEQ ID NO 5337; 1297pp + Sequence Listing; English.

XX The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a protein of the invention.
 XX

SQ Sequence 93 AA;

Alignment Scores:
 Pred. No.: 1,42e-26 Length: 93
 Score: 309.00 Matches: 57
 Percent Similarity: 76.23% Conservative: 4
 Best Local Similarity: 71.23% Mismatches: 18
 Query Match: 3.20% Indels: 1
 DB: 22 Gaps: 0

US-09-966-880A-9 (1-5514) x AAM96679 (1-93)

QY 5273 TTTTNTTTTTTTTTTTTTCAGATGGAGTTTGTGCTTGTGCTCCCATGCTGGAGTGA 5332
 Db 4 PhePhePhePhePhePhePheGluThrGluSerHisSer-ValThrGlnAlaGlyIleGl 23
 QY 5333 ATGGCATGAYCATAGTCTCACTGCAACCTCCACCTCCTGGTTCAAGCAAGCTGCGCT 5392
 Db 23 ntrpHisTyrLeuSerSerLeuGlnProProProProArgPheLysGlnPheSerCysLe 43
 QY 5393 CAGCTCCCGGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5452
 Db 43 uSerLeuLeuSerSerTrpAspTyrArgHisThrProProHisAlaAlaAsnPhcCysIl 63
 QY 5453 TTTTATGATGATGGGTTTTCACCATGTTGGCCAGCGTGGTCTCAAACTCCCTGACCTC 5510
 Db 63 ePheSerArgaspGlyValSerProCysTrpSerGlyTrpSerArgThrProAspLeu 82

RESULT 8

AAU31902
 ID AAU31902 standard; Protein; 107 AA.

XX AAU31902;

XX 18-DEC-2001 (first entry)

XX

DE Novel human secreted protein #2393.
 XX Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200179449-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 16-APR-2001; 2001WO-US08656.
 XX
 PR 18-APR-2000; 2000US-0552929.
 PR 26-JAN-2001; 2001US-0770160.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-611725/70.
 XX
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy -
 XX
 PS Claim 20; Page 524; 765pp; English.
 XX
 CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising
 CC the nucleic acids encoding the polypeptides and cells genetically
 CC engineered to express them are also useful for producing the proteins.
 CC The proteins are useful in genetic vaccination, testing and
 CC therapy, and can be used as nutritional supplements. They may be used to
 CC increase stem cell proliferation; to regulate haematopoiesis; and in
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and
 CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
 CC sequences of novel human secreted proteins of the invention.
 XX
 SQ Sequence 107 AA;

Alignment Scores:
 Pred. No.: 2e-26 Length: 107
 Score: 308.00 Matches: 60
 Percent Similarity: 77.50% Conservative: 2
 Best Local Similarity: 75.00% Mismatches: 17
 Query Match: 3.19% Indels: 1
 DB: 22 Gaps: 0

US-09-966-880A-9 (1-5514) x AAU31902 (1-107)

QY 5274 TTTTNTTTTTTTTTTTTTCAGATGGAGTTTGTGCTTGTGCTCCCATGCTGGAGTGA 5333
 Db 3 PhePhePhePhePhePhePheGluThrLeuValAlaGlnAlaGlyVal*** 22

QY 5334 TGGCATGAYCATAGTCTCACTGCAACCTCCACCTCCTGGTTCAAGCAAGCTGCGCTC 5393
 Db 23 LeuHisAspProGlu***LeuGlnProProProProGlyPheLysArgPheSerCysLeu 42

QY 5394 AGCTCCCGGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5453
 Db 43 SerLeuProSerSerTrpAspTyrArgHisThrProProProProAlaAsn-Phc***Ph 62

QY 5454 TTTTATGATGATGGGTTTTCACCATGTTGGCCAGCGTGGTCTCAAACTCCCTGACCTCA 5511
 Db 62 eLeuValGluThrGlyPheHisHisValGlyGlnAlaGlyLeuGluLeuLeuThrSer 81

RESULT 9
ABP41552
ID ABP41552 standard; Protein: 110 AA.
XX
AC ABP41552;
XX
DT 22-AUG-2002 (first entry)
XX
DE Human ovarian antigen HSPSI60, SEQ ID NO:2684.
XX
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive.
XX
OS Homo sapiens.
XX
PN WO200200677-A1.
XX
PD 03-JAN-2002.
XX
PF 07-JUN-2001; 2001WO-US18569.
XX
PR 07-JUN-2000; 2000US-209467P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
DR WPI; 2002-147878/19.
XX
DR N-PSDB; ABQ54629.
XX
PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
PT ovarian cancer), immune disorders, cardiovascular disorders and
PT neurological diseases -
XX
PS Claim 11; SEQ ID NO 2684; 2922pp; English.
XX
CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents a human ovarian antigen of the invention.
CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 110 AA;

Alignment Scores:

Pred. No.:	2.66e-26	Length:	110
Score:	307.00	Matches:	72
Percent Similarity:	72.48%	Conservative:	7
Best Local Similarity:	66.06%	Mismatches:	22
Query Match:	3.19%	Indels:	9
DB:	23	Gaps:	3

US-09-966-880A-9 (1-5514) x ABP41552 (1-110)

QY	3865	TTCCATTATTTTAAATTTT-----CTTCTTATTTTTTGAGGCAAGTCTG	3818
DB	3	PhePhePhePhePheLeuPheSerAspValIleLeuTyrLeuPheLeuArgGlnAsn---	21
QY	3817	GTCCTGTCAACCCAGGCTGGAGTGCAGTGGCGCAAGCACTGCTCAGTCAGCAGCCTTGACCTC	3758
DB	22	---LeuSerProArgLeuGluCysSerGlyMetIleSerAlaHisCysArgLeuTyrLeu	40
QY	3757	CTGGGCTCCAAAGATTCTCTGCTCAGCCCTTA-GTAGCTGGGACACAGCAGCATGC	3699
DB	41	MetGlySerSerAspSerProAlaSerAlaSerGlnValValGlySerThrGlyThrHis	60
QY	3698	CACCATGCCCGCTAATTTTGT-----ATAGAAACAGGGTTTGGCATGCTCTCAG	3645
DB	61	HisHisThrGlnLeuIlePheValPheLeuValGluMetGlyPheHisValAspGln	80
QY	3644	GCTGCTTGAACCTCCAGGCTCAAGCGATCCACCACTCAGCCTCCAGAGTGCTGG	3585
DB	81	AlaGlyLeuGluLeuLeuThrSerSerAspProThrSerAlaSerGlnSerAlaGly	100
QY	3584	ATTACAGCGTGAGCCACCACTGG 3558	
DB	101	IleThrGlyValSerHisCysThrTrp 109	

RESULT 10

AAU31725
ID AAU31725 standard; Protein: 175 AA.

XX AAU31725;

XX 18-DEC-2001 (first entry)

XX Novel human secreted protein #2216.

XX Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX Homo sapiens.

XX WO200179449-A2.

XX 25-OCT-2001.

XX 16-APR-2001; 2001WO-US08656.

XX 18-APR-2000; 2000US-0552929.

XX 26-JAN-2001; 2001US-0770160.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-611725/70.

XX Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -

XX

DB: 22 Gaps: 0

US-09-966-880A-9 (1-5514) x ABB11464 (1-241)

QY 5264 CTTTCCTTTT...TTTTCAGATGGAGTTTGGTCTTGTGGCCATGC 5323

Db 145 LeuHisPheCysPhePheGlyPhePheGluMetGluSerCysSer-ValAlaGlnAl 164

QY 5324 TGGAGTGAATGGCATGAYCATGCTCACTCAACCTCCTCTGGTTCAAGCAAAG 5383

Db 164 aGluValGlnTrpHisAspLeuArgSerLeuGlnAlaProProGlyPheThrProPh 184

QY 5384 CTGTCGCTCAGCTCCGGGTAGATGGATTACAGCGCCACACACACTCGGCTAA 5443

Db 184 eSerCysLeuSerLeuProGlySerTrpAspTyrArgArgProProArgProAlaAs 204

QY 5444 TGTTCGTATTTTAGTAGAGATGGGTTTACCATTGTTGGCCAGGCTGCTCAAACTCC 5503

Db 204 nPheCysIlePheSerArgAspGlyValSerProCys***ProGlyTrpSerArgSerPr 224

QY 5504 TGACCTC 5510

Db 224 oAspLeu 226

RESULT 12

AAB75337

ID AAB75337 standard; Protein; 76 AA.

XX AAB75337;

AC AAB75337;

XX 03-APR-2001 (first entry)

DE Human secreted protein sequence encoded by gene 47 SEQ ID NO:156.

XX Human; immunosuppressive; antiarthritic; antirheumatic; nootropic;

KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;

KW neuroprotective; antibacterial; virucide; fungicide; opthalmological;

KW vulnery; autoimmune disease; hyperproliferative disorder; cancer;

KW cardiovascular disorder; cerebrovascular disorder; infection;

KW nervous system disorder; ocular disorder; chemotaxis; food additive;

KW secreted protein.

XX Homo sapiens.

OS WO200077021-A1.

PN 21-DEC-2000.

XX 01-JUN-2000; 2000WO-US15135.

PF 11-JUN-1999; 99US-0138632.

PR (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

XX Rosen CA, Ruben SM, Komatsoulis GA;

PI WPI; 2001-071257/08.

DR N-PSDB; AAF63835.

XX Nucleic acid molecules encoding human secreted proteins, used in

PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and

PT Parkinson's diseases and cancers -

XX Disclosure; Page 104; 530pp; English.

XX This invention relates to polynucleotide sequences AAF63789 - AAF63836

CC which encode human secreted proteins AAB75260 - AAB75287. Included in the

CC invention are protein sequences AAB75288 - AAB75341 which are fragments

CC of the secreted proteins and amino acid sequences with which these

CC fragments share homology. Examples of the activities of the proteins and

CC polynucleotides and the activities of their agonists and antagonists

CC include, immunosuppressive; antiarthritic; antirheumatic;

CC antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;

CC nootropic; neuroprotective; antibacterial; virucide; fungicide;

CC opthalmological; and vulnerary activity. The protein and polynucleotide

CC sequences, their agonists and antagonists may be useful for treating,

CC preventing and diagnosing diseases and disorders such as autoimmune

CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders

CC e.g. neoplasms of the breast or liver, cardiovascular disorders

CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,

CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,

CC infections caused by bacteria, viruses and fungi and ocular disorders

CC e.g. corneal infection. The polypeptides can also be used to aid wound

CC healing and epithelial cell proliferation, to prevent skin aging due to

CC sunburn, to maintain organs before transplantation, for supporting cell

CC culture of primary tissues, to regenerate tissues and in chemotaxis. The

CC polypeptides can also be used as a food additive or preservative to

CC increase or decrease storage capabilities. Included in the invention are

CC oligonucleotides AAF63780 - AAF63788 and peptide AAB75239 which are used

CC in the identification and characterisation of the DNA and protein

CC sequences of the invention.

XX SQ Sequence 76 AA;

Alignment Scores:

Pred. No.: 4.93e-26 Length: 76

Score: 304.00 Matches: 56

Percent Similarity: 74.03% Conservative: 1

Best Local Similarity: 72.73% Mismatches: 19

Query Match: 3.15% Indels: 1

DB: 22 Gaps: 0

US-09-966-880A-9 (1-5514) x AAB75337 (1-76)

QY 5282 TTTT...TTTTCAGATGGAGTTTGGTCTTCTGCCATGCTGGAGTGAATGGCATCA 5341

Db 1 PhePhePheGluMetGluSerHisSer-ValThrGlnAlaGlyValGlnTrpHisAs 20

QY 5342 YCATAGCTCACTGCAACCTCCACCTCCCTGGTTCAGGAAAGCTGTCGCTCAGCCTCCC 5401

Db 20 pleuclySerLeuGlnProProProProArgPheThrProPheSerCysLeuSerLeupr 40

QY 5402 GGTAGATGGGATTACAGCGCCGCCACACACACTCGGCTAATGTTTGTATTATTAGTAG 5461

Db 40 oSerSerTrpAspTyrArgGlnProProArgProAlaAsnPheCysIlePheGlyAr 60

QY 5462 AGATGGGTTCACCATGTTGGCCAGGCTGCTCAAACTCCTGACCTC 5510

Db 60 gaspGlyValSerProCysTrpProGlyTrpSerGlnThrProAspLeu 76

RESULT 13

AAR95913

ID AAR95913 standard; Protein; 397 AA.

XX AAR95913;

AC AAR95913;

XX 13-NOV-1996 (first entry)

DT Neural thread protein.

DE Neural thread protein; NTP; diagnosis; detection;

XX Alzheimer's disease; neuroectodermal tumour; malignant astrocytoma;

KW monoclonal antibody; binding fragment.

XX Homo sapiens.

OS WO9615272-A1.

PN 23-MAY-1996.

XX 14-NOV-1995; 95WO-US17111.

PD 14-NOV-1994; 94US-0340426.

PR (GEHO) GEN HOSPITAL CORP.

XX
PI De LA MONTE S, Wands JR;
XX WPI; 1996-259865/26.
DR N-PSDB; AAT27738.
XX
XX Detection of neural thread protein in diagnosis of Alzheimer's
PT disease - also NTP DNA and protein sequences used in gene and
PT anti:sense therapy
XX
PS Claim 22; Page 171-172; 238pp; English.
XX
XX A method for detecting the presence of neural thread protein (NTP)
CC having a molecular weight of 8, 14, 17, 21, 26 or 42 kD in a human
CC subject comprises (a) contacting a sample from a human subject that
CC is suspected of containing the NTP with at least one molecule
CC capable of binding to the protein; and (b) detecting any of the
CC molecule bound to the protein. The binding molecule is selected
CC from an antibody free of natural impurities, a monoclonal antibody
CC or a binding fragment of either of these. The method may be used for
CC diagnosing the presence of Alzheimer's disease, neuroectodermal
CC tumours and a malignant astrocytoma in a human.
XX
SQ Sequence 397 AA;

Alignment Scores:
Pred. No.: 1,16e-25 Length: 397
Score: 304.00 Matches: 123
Percent Similarity: 44.51% Conservative: 31
Best Local Similarity: 35.55% Mismatches: 114
Query Match: 3.16% Indels: 80
DB: 17 Gaps: 13

US-09-966-880A-9 (1-5514) x AAR95913 (1-397)

QY 4441 CCTTCCCTCCACAGGACAACTGATTGCTTCTTTTCC----- 4394
DB 77 ProSerArgProLysValLeuAspThrGlyLeuSerThrMetProGlyLeuCysLeu 96
QY 4393 -----ATTTTTGGAGATGAGTTTCACTGTGTGGCCAGGCTTGAACCTCT 4340
DB 97 AlaAsnPheCysGlyArgAsnArgValSerLeuMetCysProSerTrpSer-----Pro 114
QY 4339 GAGTTCTAGCAATCTCCCA-CGTGGCTCTCAAGTGTGGGATACAGGTGTGAGCT 4281
DB 115 GluLeuLysGlnSerThrCysLeuSerLeuProLysCysTrpAspTrpArg--AlaA 134
QY 4280 CTATGCTGCCCTGATTTGCTTCTCTAACTA-----TAAGTTACTTTCTTCT 4230
DB 134 laValProGlyLeuPheLeuPheLeuPheLeuArgHisArgCysProThrLeuThrGlnA 154
QY 4229 AGAATTTCTGATGG-----CATACATTATCTCTCTTCATCTGCTGACTCTTTCATTTA 4176
DB 154 spGluValGlnTrpCysAspHisSerSerLeuGlnProSerThrLeuArgSerSerIle- 173
QY 4175 GCCTCATGACTTTGGGATTCATCCCAAGTACTCTTCTTGTACTGCAGGGGGTAT 4116
DB 174 -----LeuLeuProGlnProProLysValAla- 182
QY 4115 TCATTGTCATGATAGACATTTTGTATTATTTGCTTCTTGTGATGGATTTGGAT 4056
DB 183 --GlyThrLysAspMetHisHisThrTrpLeuIle-----PheIlePhe--- 197
QY 4055 AGTCTCTACTTTGGGTTTGTG----- 4033
DB 198 -----IlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGlyValG 215
QY 4032 -----AATAAGGTACTAGAACATTCAGTACAGTCTTT----- 3997
DB 215 lnTrpArgAsnLeuGlySerLeuGlnProLeuProProGlyPheLysLeuPheSerCysP 235
QY 3996 -----GTGTGGATAT 3987
DB 3996 -----

Db 235 roSerLeuLeuSerSerTrpAspTyrArgArgProProArgLeuAlaAsnPhePheValP 255
QY 3986 ACATTGAGAAATTGATATTTTAAATATGTAGTGCAGGTGCACAGACCTTGACTTTTAAAGTT 3927
Db 255 heLeuValGluMetGlyPheHisValArgGlnValAspAlaArgSerLeuAspLeuV 275
QY 3926 AATATACCAAAATATCAGCCTAACTAAGTAGTACTCATCTAGGACACAGCTCTTTGTAGTTG 3867
Db 275 alileCysLeuProArgProLysValLeuGlyLeu---GlnAspValThrProThrA 294
QY 3866 TTTCATTTTAAATTTTCTTCTTATTTTTCAGCAAGGTCTGTCTCTGTCCACC 3807
Db 294 laArgPro-IlePheAsnPheCys-----LeuPheGluMetGluSerHisSerValThr 311
QY 3806 CAGCTGGAGTGCAGTGGCCCAAGCACTGCTCAGTGCAGCTTTCAGCTCTCTGGCTCCCA 3747
Db 312 GlnAlaGlyValGlnTrpProAsnLeuGlySerLeuGlnProLeuProGlyLeuLys 331
QY 3746 AGATTCTCTCCGCTCAGCC-CCTAGTACCTGGACACAGGACATGCCATGCCCG 3688
Db 332 ArgPheSerCysLeuSerLeuProSerSerTrpAspTyrGlyHisLeu-HisHisThrPr 351
QY 3687 GCTAATTTTGT-----ATAGAAACAGGGTTTTCCTGCTGCTGCTGCTTCA 3634
Db 351 oLeuIlePheValPheSerLeuGluAlaGlyPheHisHisIleCysGlnAlaGlyLeuLys 371
QY 3633 ACTCCAGGCTCAAGCGATCCACCTCAGCTCCAGCTCCAGTGTGGGATTTACAGGCT 3574
Db 371 sleuLeuThrSerGlyAspProProAlaSerAlaPheGlnSerAlaGlyIleThrGlyVa 391
QY 3573 GAGCCACAC 3564
Db 391 lThrProHis 394

RESULT 14
ABG12796
ID ABG12796 standard; Protein; 211 AA.
XX
AC ABG12796;
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #12787.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS76983.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID NO 43155; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and

A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of the

Alignment Scores:

Pred. No.: 9.59e-38 Length: 673
Score: 445.00 Matches: 180
Percent Similarity: 32.32% Conservative: 44
Best Local Similarity: 25.97% Mismatches: 161
Query Match: 4.62% Indels: 309
DB: 4 Gaps: 22

US-09-966-880A-9 (1-5514) x F40201 (1-673)

```
QY 5510 GAGGTCAGGAGTTTGACACAGCCTGGCCCAACATGGTGAACCCCATCTCTACTATAAAT 5451
D 5511 |||||
D 248 GluValArgSerLeuArgProAla***SerThrTrp***AsnCysValSerIleLysAsn 267
QY 5450 ACRAACATTAGCCGAGTGTGGTGGCGCCTGTAATCCCATCTACCCGGGAGGCTGAG 5391
D 5451 |||||
D 268 Thr***IleSerGlnGluTrpArgProMetProValIleProAlaThrArgGluThrGlu 287
QY 5390 GCGACAGCTTTGGTTGAACCCAGGAGGTGGAGGTTCGAGTCAGTCATGTCATGCCATC 5331
D 5331 |||||
D 288 AlaglyGluSerLeuGluProArgArgLeuArgLeuGln***Ala-LysThrCysHisCy 307
QY 5330 CACTCCAGCATGGGCACACAGACCAAACTCCATCTCAAAAAA*****AAAAA 5271
D 5271 |||||
D 307 sthrProAla-GlyValSerGlnGlyProIleSerLysThrLysThrLysGlnLysL 327
QY 5270 AAGAAAGAGTAATACTTGTGATCATTTGTCAAACCTTGAATACTAGCAAAATAGAGACTC 5211
D 5211 |||||
D 327 yslsLys-----LysAsnProGlnLys- 334
QY 5210 TAAATTTTAAATATTAAGATTTTTAAATCTCAAAATATCATAGTCTCTAATAATA 5151
D 5151 |||||
D 334 ----- 334
QY 5150 ATTAGACTTAGTGTCTAATGACAAATGACTGTTAAGATTCTTGCTTTTATTCCTCTAGTCT 5091
D 5091 |||||
D 334 ----- 334
QY 5090 TTCTACAGCACATATTTACATGATTGG-----ATTATACTGATTGCATGTGATT 5037
D 5037 |||||
D 335 -----Gln*****PheCysPheTrpGlyPhePhePhePheCysPheValP 353
QY 5036 TGTACTCTGCTGCTTTTAACTAGTCAAAATGATCTAAATATCTTGTCTGTTGTTCCC 4977
D 4977 |||||
D 353 heVal----- 354
QY 4976 ACATAATGACTGATTTCTCCTCTCTTGTGTTTCTGCTCCTAGCACAGTGTCTTGCA 4917
D 4917 |||||
D 354 ----- 354
QY 4916 TCATAAGCACTAATTTGTAAGCATTTGTGSAATACTCTGGTAACCAAGTTTCCCTAA 4857
D 4857 |||||
D 354 ----- 354
QY 4856 AGAAATCAAGAGTCCCATTTGCTCCTCTCTTGTGTTTCTGCTCCTAGCACAGTGTCTTG 4797
D 4797 |||||
D 355 -----PheGluIleGlyProCysSerAspThrProAlaGlyValGlnTrpGlnV 371
QY 4796 TATGAAGACACAGTGTGCTGCTGATTTTGTCCCTAGTCTTTTCCAGAAATGCAATTTATC 4737
D 4737 |||||
D 371 alLeuAlaHis-----CysSerLeuA 378
QY 4736 CAATTTGTGAGCTAGTGTCTACCTAATGGTACGGGGCAGAGTGCAGAGAAATTTGG 4677
D 4677 |||||
D 378 snLeuGlySerSerAspSerPro-----AlaSerValSerArgValAlaG 394
QY 4676 AAGTAAAGGGCTCAGGAATTTGAATAACGTTTGTGCGCAAGAGAAATTTGAACTTTCTTA 4617
```

```
394 lyIleThrGlyMetGly----- 399
4516 CACATTATTTTCTTACATTTAGGTATATTTTACATACAGTAACCTCCACCCATTAAAG 4557
399 ----- 399
4556 ATGCATAATTTCTAGCAATATTTTGGATGTAAACACACCCCAAACTGCTTACCACAGTCAGGA 4497
400 -----ArgHisSerTrp----- 403
4496 TACAGAACCTGAATCACCCCTCACCCACAGAGGCTTCTTCCCTCTTTGTCAGTCCTTC 4437
403 ----- 403
4436 CCCTCCCAACAGCAACCACTGATTTGCTTTCTTTTCCCATTTTGTGGAGAAATGG 4377
404 -----LeuIleTrpValPheLeuIleGluThrG 413
4376 AGTTTCACTGTGTGGCCAGGCTGTGAACCTCCTGAGTTCTAGCAATCTTCCACAGT 4317
413 lnPheHisValAspGlnAlaGlyLeuLysLeuThrSerSerAspLeuProSerT 433
4316 GGCTCTCTCAAAGTCTGGGATTACAGGTGTGAGCTCTATGCTGCTGCTGATTTGCTTTC 4257
433 rpSerProLysValLeuGly***GlnAla***AlaThrThrProSer***** 449
4256 TGTAACTATAAGTTACTTTTCATTTTCTAGAAATTTTCGTATGGCATACATATCTCTCCTT 4197
450 -----PheValPheGlyGlyPhePhePhePhePheAla----- 463
4196 CATGCTGCTGCTTTCATTTAGCGTCATGACTTTGGGATTCATCCCAAGTAGTCTTTC 4137
464 -----LeuPheLeuPheLeuArg***AlaLeuAlaLeuThrProArg----- 477
4136 CTTTGTATGTCAGGCTGATTTCCATTTGCATTTGATGATATAGCACATTTTGTTTTGTTC 4077
478 --LeuGluCysSerGlyLysPhe----- 484
4076 CTTGTTGATGATATTTGGATAGTCTCTACTTTGGGGTTTGTGTAATAAGGTACTATGA 4017
485 -----TrpLeuThrAlaAlaSerThrSerTrp-----ValGlnA 496
4016 ACATTACGTCACAGTCTTT-----GTCTGGATATACATTTGAGA 3978
496 laIleLeuLeuProLeuSerProVal***LeuGlyLeuGlnAlaTrpAlaAlaIleProG 516
3977 AT-----TGATATTTT-----A 3966
516 ly***PheMetTyrPhe*****ArgHisSerPheThrMetLeuIleArgLeuValSerA 536
3965 ATAATGTAGTCAGGTGACAGACCCCTGACTTTAAAGTTAAATACCAAAA----- 3916
536 snSer***ProGlnValIleCysProGly-----LeuProLysCysTrpAspA 553
3915 -----TATCAGCCTAACTAAGTAGTCTCATCTAGGACAAACAGCTCTTTGTAGTCTTTTCCA 3861
553 spArgArgGluProProHisProAla-----*****LeuPheLeuGlyV 569
3860 TTTATTTTAAATTTTCTTCTTATTTTGTAGCAAGGTCTGCTCTGT----- 3811
569 alPhePhePhePhePheLeuLeuCysPhe-----CysPhe***AspA 583
3810 -----CACCCAGGCTGGAGTGCAGTGGCGCAAGCAGCTGCTAGTCAGCAGCCT 3765
583 rgProLeuLeu***HisProGlyTrpSerAlaValAla-SerPheGlySerLeuGlnPro 602
3764 TGACCTCTGGCTCCAAAGATTTCTGCTGCTCAGCCC-CCTAGTACTGGGACACACAGG 3706
603 GlnProProGlyPheLysArgPheSerCysLeuProCysSerTrpAspTrpArg 622
3705 CACATGTCACCATCCCGGCTAATTTTGTGTTATAGAAACAGG-----GTTTTGCGCATGT 3652
```

Qy	3993	ACACAAAGACTGGTACGTGAATGTCATAGCTATCTTTATTACACAAACCCCAAGTAGAG	4052
	:		:
Db	157	sn***ProGlyMetAlaAlaHisAlaCysAsnProSerTyrThrGlyAspArgGlyArgA	177
	:		:
Qy	4053	ACTATCCAAATATCCCATCAACAGTGAACAAATAACACAAATGTGCTATATCC-----A	4106
	:		:
Db	177	rgIleAla***ThrGlnGlu-----ValGluAlaAlaValSerGlnAsnL	192
	:		:
Qy	4107	TGCAATGGAATACACCCCTGCAGTACAAAGGAAGAAGCTACTTCTGGGGATCAATCCCAAG	4166
	:		:
Db	192	euProLeuHisSerSerArgGlyValArgAlaArgAlaTyrLeuLysAsnLysA	212
Qy	4167	TCATGAGCGCTAAATGAAAGAGTCAGACATGAAGGAGGAGATAATGTATGCATACCAAT	4226
Db	212	lalys-----	213
Qy	4227	TCTAGAAAATGAAAGTAACTTATAGTTACAGAAGACNAATCAGGGCAGGCATAGAGCTC	4286
	:		:
Db	214	-----LysLysLysLysLysProProLysThrLys*****LeuGlyValValAla	232
Qy	4287	ACACCTCTAATCCAGCACCTTTGAGAGGCCACGTGGGAAGATTGCTAGAACTCAGAGTT	4346
	:		:
Db	232	isAlaCysHisProSerThrLeuGlyAspGlnGluGlyArgSerLeuGluValArgSerL	252
	:		:
Qy	4347	CAAGACACGCTGGGCCAACACAGTGAACCTCCATTCTCCACAAAAATGGGAAAAAAGAA	4406
	:		:
Db	252	euArgProAla***SerThrTrp***AsnCysValSerIleLysAsnThr**IleSerG	272
	:		:
Qy	4407	AGCAATCAGTGGTGTCTCTGTG-----GGAGGGGAA	4439

Db 2/2 *ingluirfPArgProMetCProValIleProAlaIthrArgGluIthrGluAlaGlyCysSerL* 252
 4440 *GGACTGCAAGAGGGAAGACGCTCGTGGGTGAGGGTGGTGATTCAGCTTCTGTATCC* 4499
 Db 292 *eugluPProArgArgLeuArgLeuGln***Ala* 307
 4500 *TGACTGTGGTAGCAGATTGGGGTGTTTACATCCAAAATATTCGTAGAATTATGCATCTT* 4559
 Db 307 *ythrProAlaGlyVal-----SerG* 314

	QY	4560	AAATGGGTGAGTTTACTGTATGTAATAATTATCAATCTAGAAGAAAATAAATGTGTAA	4561
	Db	314	luGlnGlyProIle-SerLysThrLysThrLysGlnLysLysLysAsnProGln	333
	QY	4620	GAAAGGTTTCATTTCTCTCCACGAAACGTTTCTCAAATCTCTGAGCCCTTACTTCGC	4679
	Db	334	LysGln*****PheCysPheTrpGlyPhePhe-----	345
	QY	4680	AAATTCTCGCACTTCTGCCCGTA-----CCATTAGTGACAGCAC	4721
	Db	346	--Phe-PhePheCysPheValPheValPheGluIleGlyProCysSerAspThr	364
	QY	4722	TAGCTCCACAAATGG--- ::: 	4737
	Db	364	oAlaGlyValGlnTrpGlnValLeuAlaHisCysSerLeuAsnLeuLeuClySerSerAs	384
	QY	4737	-----	4737
	Db	384	pSerProAlaSerValSerArgValAlaGlyIleThrGlyMetGlyArgHisSerTrpLe	404
	QY	4738	-ATAATGCATTCTTGGAAAGACTAGG-----GACAAATCCAGCATCACTTGTGCT	4790
	Db	404	uileTyrrValPheLeuIleGluThrGlnPheHisValAspGlnAlaGlyLeuLysLe	424
	QY	4791	T-----TCATATCAACCACGCTGTACAGCTTGTCTCTCTGCTGCTG	4829
	Db	424	uLeuThrSerSerAspLeuProSerTrpSerProLysValLeuGly**cInAla***Al	444
	QY	4830	CAGCTGCAATGGGACTCTTTGATTCTTTTAAGGAACACTTGGGTTAC-----	4875
	Db	444	aThrThrProSer*****PheValPheGlyGlyPhePhePhePheAlaLe	464

QY 4875 ----- 4875
Db uPheLeuPheLeuArg**AlaLeuAlaLeuThrProArgLeuGluCysSerGlyLysPh 484
QY 4876 -----CAGAGTATTTCCACAATCTATTCAATTAGTGTATGATATGCAAGACAC 4928
Db eTrpLeuThrAlaAlaSerThrSerTrpValGlnAlaIleLeuLeuProLeuSerProVa 504
QY 4929 TGTGCTAGGAGCCGCAAAACAAAGAGGA-----GGAGAAATCAGTCATTATGTGGGAAC 4982
Db 504 l***LeuGlyLeuGlnAlaTrpAlaAlaIleProGly***PheMetTyPhe*****Ar 524
QY 4983 ACATAGCAAGATATTAGATCATTTTACTAGTTTAAAAAGCAGCAGTAGTACAAAATCA 5042
Db 524 gHisSerPheThrMetLeuIleArgLeuValSerAsnSer***ProGlnVal----- 541
QY 5043 CACATGCATCAGTATAATCCAAATCATGTAATATGTGCTGTAGAAAGAGTAGAGGAA 5102
Db 542 -IleCys-----ProGlyLeuProLysCysTrp-----AspAspArgAr 555
QY 5103 TAAACACAAGAAATCTTAAACAGTCATTGTCTATTAGACACTAAGTCTAATTATTATTAG 5162
Db 555 gGlu----- 556
QY 5163 ACATATGATATTGTAGATTTAAAAAATCTTTAATATTTTAAAAATTTAGAGCTCTTCTAT 5222
Db 556 ----- 556
QY 5223 TTTTCCATGATTCAAGTTTGACATGATCAAGTATTACTCTTTCTTTTCTTTTCTTTT 5282
Db 557 ----ProProHisProAla*****LeuPheLeuGlyValPhePhePhePheLe 575
QY 5283 T-----TTTTTTTTTGATGGAGTGTGTTGGTCTGTTGCCCATGCTGGAGTGAATGG 5336
Db 575 uleuCysPheCysPhe***AspArgProLeuLeu***HisProGlyTrpSerAlaValA 595
QY 5337 CATGAYCATAGCTACTGCAACCTCCACCTCGGGTTCAAGCAAAAGCTGTCCGCTCAGC 5396
Db 595 la-SerPheCysSerLeuGlnProGlnProGlyPheLysArgPheSerCysLeuCys 614
QY 5397 CTCCTCCGGTAGATGGATTACAGCGCCGCCACACACACCTCGGCTAATGTTGTATTTT 5456
Db 615 LeuProCysSerTrpAspTyArgHisGlyProPheLeuAlaAsnLeuCysIlePhe 634
QY 5457 AGTAGAGATGGGTTTCACCATCTGGCCAGGCTGTCTCAAACTCCTGACCTCAGA 5513
Db 635 AsnArgAspThrValSerProCys***SerGlyTrpSerGlnThrProAspLeuLys 653

RESULT 4

C40201
artifact-warning sequence (translated ALU class C) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: C40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: C40201
A:Molecule type: DNA
A:Residues: 1-613 <CLA>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
A:Reference number: A40200; PMID:92241891; PMID:1572661
C:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
in-frame stop codons are shown as 'x'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:

Pred. No.: 8.64e-34 Length: 613
Score: 407.50 Matches: 181

Percent Similarity: 33.92% Conservative: 52
Best Local Similarity: 26.35% Mismatches: 221
Query Match: 4.23% Indels: 234
DB: 4 Gaps: 17
US-09-966-880a-9 (1-5514) x C40201 (1-613)
QY 5510 GAGGTCAGAGTGTGAGACAGCTGGCCAAACATGGTGAACCCCATCTCTACTAAAAAT 5451
Db 125 GluAlaArgSerSerArgProAlaTrpProThrTrpGlnAsnProIleSerThrLysAsn 144
QY 5450 ACAACATTAGCCGAGTGTGGTGGCGCCTGTAAATCCATCTACCCGGGAGGTGAG 5391
Db 145 ThrLysAsn***LeuGlyMetValArgAlaProValValProArgThrGlnLysAlaGlu 164
QY 5390 GGCACAGCTTCTTCAACCCAGGAGGTGGAGGTTCAGTGAGCTATGTCATGCCATTC 5331
Db 165 AlaGlyGluTrpHisLysProGlyArgTrpSerLeuGln***AlaGluIleSerProLeu 184
QY 5330 CACTCCAGCATGGGCAACAGACCAAACTCCATCTCAAAAAAATAAAAAAATAAAAA 5271
Db 185 HisSerSerLeuGlyAsp-ArgValArgLeuCysLeuLysLysLysLysLysLys** 204
QY 5270 AAGAAAGATTAATCTGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 5211
Db 204 *****ArgGlyAlaValAla----- 211
QY 5210 TAAATTTTAAATATTTAAAGATTTTAAATCTCAATATCATAGTGTCTAATAATA 5151
Db 211 ----- 211
QY 5150 ATTAGACTTAGTCTCTAATGACAATGACTGTTAAGATTTCTGTGTTTATTTCCTCTAGTCT 5091
Db 212 -----HisAlaCysAsnProSerThrLeuG 220
QY 5090 TTCTACAGGCACATATTTACATGATTGGATTACTGATTGCTGATGCTGATTGTTGACT 5031
Db 220 yGlyLysGlyGlyTrpIleMetArgProGlyValArgAsp----- 233
QY 5030 CTGCTCTTTTAACTAGTCAAAATGATCTAAATATCTTGCTATGTTGTTCCACATAA 4971
Db 233 ----- 233
QY 4970 TGACTGATTTCT 4911
Db 233 ----- 233
QY 4910 GCCTAATTTGAATAGCATTTTGGAAATACTCTGTTAACCCTAAGTTTCCTTTAAAGAAAT 4851
Db 234 -----GlnProGlyGlnHisGlyLysThrProPheLeuLeuLysIleG 248
QY 4850 CAAGAGTCCCATTTGCGAGCTGCAGACAGCAACACAGCTGTACAGCTGGTGTGATATGAA 4791
Db 248 nLysIleSerTrpAlaTrpCys-----GlyArgLeu***SerHisValArgArgAr 265
QY 4790 AGCACAA-----GTGATGCTGGATTGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4746
Db 265 gLeuArgGlnGluAsnGlyIleAsnProGlyGlyGlyAlaCysSerGluProArgSerAr 285
QY 4745 GCATTTATCCAAATTTGTGGAGCTAGTCTGCTACCTAATGTTAGGGGCAAGAGTCAGA 4686
Db 285 gHisCysThrProAlaTrp----- 291
QY 4685 GAATTTGCCGAAGTAAAGGCTCAGGAATTTGAATAACGTTTGTGTCGCAAGAGAAATGAAA 4626
Db 292 -----ValThrGlu***AspSerValSerLysLysLysLysLys 305
QY 4625 CTTTTCTTACACATTTATTTTCTTACATTGAGGTATTAATTTTACATACATCAAACTCCAC 4566
Db 305 s*****PhePhePhePhePhePhePhePhePhePhePhePhePhePhePhePhePhe 325
QY 4565 CCATTTAAGATGCATAATTTCTAGCAATATTTTGGATGTAAACACCCCAACTGCTACCA 4506
Db 4565 ----- 4506

```
Db 325 aglyVal----- 327
QY 4505 CAGTCAGATACAGAACCTGAATCACCACCCTCACCCACCAGACGCTTCTCCCTCTTG 4446
Db 328 ----GlnTrpArgAspLeuGlySerLeuGlnAlaProProGlyPheMetProPheSe 346
QY 4445 CAGTCCTTCC-----CCTCCACACAGGACAAACCACTGATT 4410
Db 346 rCysLeuSerLeuLeuArgThrTrpAspTyrArgArgProHisHisAlaGlnLeu----- 364
QY 4409 GCTTTCTTTTTCCTCCATTTTGTGGAGAAAGTTCACATGCTTGTGCCAGGCTGTGC 4350
Db 365 -----IlePheCysIlePheSerArgAsnGlyValLeuProCysTrpProGlyTrpSe 382
QY 4349 TTGAACCTCTGAGTTCTAGCAATCTCCACAGTGGCCTCTCAACGTCGAGGATACAGG 4290
Db 382 rArgThrPro-GlyLeuMetIleHisProProLeuProProLysValLeuGlyLeuGlnA 402
QY 4289 TGTGAGCCTCTATGCTGCGCTGATTGCTTCTGTAACATATAAGTTACTTTTCATTTTCT 4230
Db 402 Ia**AlaThrAlaProArgPhePhe----- 410
QY 4229 AGAATTTGATGGCATACATTATCTCTCTTCATGCTGACTCTTTCATTTACGTCGA 4170
Db 411 -----PhePhePhePhe*****LeuArgGlnSerLeuThrLeuSerProArgLeuG 429
QY 4169 TGACTTTGGGATTTCATCCCAAGTAGCTTCTCTTTGTACGTCAGGTTGGTATTCCTATT 4110
Db 429 luCysAsnGlyGluIle-----SerAlaHisCysLysLeu-----HisL 442
QY 4109 GCATGGATATACGACATTTTGTATTATTGTTTCTGATGTTGATGATATTGGATAGTCTC 4050
Db 442 euProGlyLeuCysHis----- 449
QY 4049 TACTTTGGGTTTGTGTAATAAGTACTATGCAACATTCAC---GTACCAGTCTTGTGT 3993
Db 449 roAlaSerAlaPheCysValArgGlyThrThrGlyAlaArgThrMetProSer***PheP 469
QY 3992 GGATATACATTCAGAAATGATATTTTAATAATGTAGTCAGGTGACAGACCCCTGACTTTA 3933
Db 469 heValPheLeuValGluMetGlyPheCysHisValGlyGlnAla----- 483
QY 3932 AAAGTTAAATACAAATATACGCCTAATAAGTAGCTCATCTAGACAAACAGCTCTTGT 3873
Db 484 -----GlyLeuGluLeuLeuA 489
QY 3872 TAGTGTCTTCATTATTTT----- 3853
Db 489 laSer***SerThrHisLeuCysLeuProLysCysTrpAspTyrArgArgGluProLeuH 509
QY 3852 -----TAATTTTCTTCTTATTTTGTGGCAAGTCTGTGCTCTGTCACC 3807
Db 509 isProAla*****PhePhePhePhePhe***AspArgValSerLeuCysHisP 529
QY 3806 CAGGTGGAGTCAGTGGCGCAAGCACTCTCAGTCGACGCTTGACTCTCTGGCTCCAA 3747
Db 529 roGlyTrpSerAlaMetAlaArgSerArgLeuThrAlaSerSerThrSerArgValTyrA 549
QY 3746 AGATATCTGCTGCTCAGCCCTCTAGTAGCTGGGACCACAGCACATGCCACCATGCCGG 3687
Db 549 laIleLeuLeuProGlnProSerAlaTyrValGlyLeuGlnAlaProAlaProCysProA 569
QY 3686 CTAATTTTGTATTAG-----AAACAGGTTTTGCCATGTGTGCTCAGGCTGGTCTGA 3634
Db 569 laAsnPheLeuTyrPhe*****LysTrpGlyPheAlaMetLeuAlaArgLeuValIserA 589
QY 3633 ACTCCAGGCTCAAGCGATCCACCCCTCAGCTCCAGCTCCGAGTGTGGGATTTACAGGCT 3574
Db 589 snSer--TrpProHisAspProThrPheAlaSerGlnSerAlaGlyIleThrGlyVa 608
QY 3573 GAGCCACCAACCC 3561
Db 608 lSerHisCysThr 612
```

```
RESULT 5
E40201
artifact-warning sequence (translated ALU class E) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: E40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: E40201
A:Molecule type: DNA
A:Residues: 1-597 <CLA>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potentl
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames o
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of
```

```
Alignment Scores:
Pred. NO.: 6,35e-27 Length: 597
Score: 342.50 Matches: 196
Percent Similarity: 33.29% Conservative: 60
Best Local Similarity: 25.49% Mismatches: 209
Query Match: 3.55% Indels: 305
DB: 4 Gaps: 29

US-09-966-880A-9 (1-5514) x E40201 (1-597)
QY 3559 CAGGTGTGTGCTCAGCCTGTAAATCCACGACCTCTGGAGGCTGAGTGGTGG-ATC 3617
Db 2 GlnValTrp**LeuMetProValIleProAlaLeuTrpGluPheLysAlaGlyArgIle 21
QY 3618 GCTTGAGCCCTCGAGTTCGAAGACCAGCCTGACACATCGGCAAAACCCCTGTTCTTATAC 3677
Db 22 Thr**GlyTrpGluPheGluThrSerLeuThrAsnMetGluLysProHisLeuTyr-** 41
QY 3678 A-----AAAATAGCCGGCATGGTGCATGTGCTGTGTCGCCAGCTACTAGGGGGCT 3731
Db 41 *LysTyrLysIleSerTrpAlaTrpTrpArgMetProValIleProAlaThrArgGluAl 61
QY 3732 GAGGCAG---GAGAATCTTTGGAGCCAGGAGTCAAGCTGCCTGACGACGTCTGCG 3788
Db 61 aGluThrGlyGluAsnCysLeuAsnProGlySerLysValCysGlyGlu--IleVal--- 79
QY 3789 CCAGTGCCTCAGCCTGGGTGACAGGACCAGA---CCTTGCCTCAAAAATAAGAAGA 3845
Db 80 ProSerTyrSerGlyLeuGlyAsnLysSerLysThrProSerGlnLysLysLys***** 99
QY 3846 AAAATTAATAAATAATGGAACAACACTACAAAGAGCTGTTGTCTCTAGCTAGCTACTAGT 3905
Db 100 ***AlaArg----- 102
QY 3906 TAGGCTGATATTTTGGTATTATTAACCTTTAAAGTCTAGGGTCTGTCACTGCACCTACATTAT 3965
Db 102 ----- 102
QY 3966 TAAATATCAATTCATCAATGTATATCCACAAAGACTGGTACGTGAATGTTTCATAGTAC 4025
Db 103 -----CysGlySerSerCysLeu***SerGlnArg----- 112
QY 4026 CTTTATTCAAAACCCCAAGTAGAGACTATCAAAATATCCATCAACAGTGAACAAAT 4085
Db 113 ---PheGlySerSerArgArg----- 118
QY 4086 AAACAAAATGTCTATATCCATGCAATGGAATACCAACCCCTGCGAGTACAANGGAAGAGCT 4145
Db 119 -----GlyGlySerProGluValGlySerLeuArgProAla 130
```


US-09-966-880A-9 (1-5514) x A40201 (1-627)

```
QY 3570 GCTCAGCGCTGTAATCCAGCACTCTGGAGGCTGAGGTGGGTGATCGCTTGACGCCCTG 3629
Db ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
QY 5 SerHisAlaCysAsnProSerThrLeuGlyGlnGlyArgIleThr--ArgSerG 24
QY 3630 GAGTTCAAGACAGCGCTGAGCAACATGGCAAAACC-----CTGTTCTATTAACAAAAT 3683
Db || ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 24 lyAspArgAspHisProGlyTyrHisGlyGluThrProSerLeuLeuLysIleGlnLysI 44
QY 3684 TAGCCGGGCATGGTGGCATGTCCTGTGTCCTCCAGCTACTAGGGG-CTGAGGCAGGAGA 3742
Db ::::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 44 IeserArgAla***CysGlyArgLeu***SerGlnLeuLeuGlyArgLeuGlnGluA 64
QY 3743 ATCTTTGGAGCCAGGAGTCAAGGCTGCACCTGAGCAGTGCCTGGCCCACTGCACCTCCAG 3802
Db || ::::::::::| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 64 snGlyValAsnProGlyGlyAlaCysSerGluProArgSerArgHisCysThrProA 84
QY 3803 CTGGGTGACAGGACCACTTGCTGCTCAAAAAATAAGAGAAAA----- 3849
Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 84 laLeuAlaThrGluArgAspSerValSerGluLysAsnLysAsnLysLys***Lys*** 104
QY 3850 -----TTAAAAATAAATG 3862
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 104 ****AlaGlyArgGlyLeuThrProValIleProAlaLeuTrpGluAlaLysAlaGly 124
QY 3863 GAACAACATACAAAGAGCTGTGTCTCTAGATGAGCTACTAGTTAGGCTGATATTTGGT 3922
Db ||::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 124 lySerArgGlyGlnGluIle----- 130
QY 3923 ATTTAACTTTAAAGTCAGGCTCTGTCACTGCACCTACATATTTAAATATCAATTCAC 3982
Db ||::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 130 ----- 130
QY 3983 ATGTATATCCACAAAGACTGGTACGTGAATGTTTCATAGTACCTTTATTCCAAAAACC 4042
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 131 -----GluThrIleLeuAlaThrThrV 138
QY 4043 CAAAGTAGAGACTA---TCCAAATATCCATCAACAAGTGAACAAATAACAAAAATGTGC- 4098
Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 138 alLysProArgLeuTyr***LysTyrLysLysLeuAlaGlyArgSerAlaGlyAlaCys 158
QY 4099 -----TATATCCATCCATGGAATACCACCTGCGAGTACAAAGGAAGAGCTACTGTG 4150
Db ||::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 158 erProSerTyrLeuGlyGly***GlyArgArgMetala***ThrArgGluAlaGluLeu- 177
QY 4151 GGGATGAATCCCAAGCATGACGCTAAATGAAGAGTCAGACATGAAGAGGAGACATAAT 4210
Db :::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 178 -----AlaValSerArgAsp-----ArgAlaThrAlaL 187
QY 4211 GTATGCCATAGCAATTTCTAGAAATGAAAGTAAGTATAGTTACAGAAAGCAAAATCAGG 4270
Db ||::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 187 euGlnProTrpArgGlnSerGluThrProSerArgLysLysIleLysThrLysAsnLysL 207
QY 4271 GCAGGCATAGAGGC-----TCACACCTGTAAATCCAGCAGCTTTGAGAGGCCACGTG 4321
Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 207 ys*****ProGlyAlaValSerArgLeu***SerGlnHisPheGlyArgProArgA 227
QY 4322 GGAAGATTGCTAGAACTCAGGAGTTCAAGACAGCGCTGGGCAACACAGCTGAACATCCCAT 4381
Db ||::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 227 rgAlaAspHis-GluValArgArgSerArgProSerTrpLeuProArg***AsnProVal 246
QY 4382 CTCACAAAAATGGGAAAAAAGAAAGCAAAATCAGTGTGTCCTGGGGGAGGGAAG 4441
Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 247 SerThrLysAsnThrLysAsn***ProGlyValValArgAlaProVal----- 262
QY 4442 ACTGCAAGAGGAAGAAGCTCTGTGGGTGAGGCTGATTCAGGTTCTGTATCTG 4501
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 262 ----- 262
QY 4502 ACTGTGTGAGCAGTTGGGTGTTTATACATCCAAAAATATTCGTAGAAATTCATCTTAA 4561
Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 263 ---ValProAlaThrTrp----- 267
```

```
QY 4562 ATGGTGGAGTCTTACTGTATGTAAATATATACCTCAATGTAGAAAAAATAATGTGTAGA 4621
Db 267 ----- 267
QY 4622 AAAGTTTCAATTCCTTGGCCAGCAACAGTTTATTCAAATTCCTGAGCCCTTTACTTCGCAA 4681
Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 268 -----GluAlaGluAlaGlyGluTrpArgGluProGlyArgArgSer 281
QY 4682 ATT---CTCTGCACCTTCGCCCGGTACCATTTAGGTGCACAGCACTAGCTCCACAAATGGA 4738
Db :::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 282 LeuGln***AlaGluIleAlaProLeuHis----- 291
QY 4739 TAAATGCATTTCTGGAAGAGACTAGGACAAATCCAGGCATCAGCTTGTGCTTCATATC 4798
Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 292 -----SerSerLeuGlyAspArgAlaArgLeuArgLeu----- 302
QY 4799 AACACAGCTGTACAGCTTGTGTGCTGCAGCTGCAATGGGCACTCTTGATTTCTTT 4858
Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 302 ----- 302
QY 4859 AAGGAACTTGGTTACCAGAGATTTCCACAAATGCTATTCCAAATTAGTGCTTATGATA 4918
Db 302 ----- 302
QY 4919 TGCAAGACACTGTGCTAGGAGCCAGAAAAACAAAGAGAGAGAGAAATCAGTCATTATGTGG 4978
Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 303 -----GlyLysLys***LysGlnLys----- 309
QY 4979 GAACAACATAGCAAGATATTAGATCANTTTGACTAGTTTAAAAAAGCAGCAGACAGACA 5038
Db :::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 310 -----IleLysLys***** 314
QY 5039 ATCACACATGCAATCAGTATATCCAAATCATGTAATATGCGCTGTAGAAAGACTAGA 5098
Db 314 ----- 314
QY 5099 GGAATAAACACAAAGAATCTTAAACAGTCATTGTCTATTAGACACTAAGCTTAATTATTA 5158
Db 314 ----- 314
QY 5159 TTAGACACTATGATATTTCAGATTTTAAAAAATCTTTAATATTATTTAAATTTAGAGCTCT 5218
Db 314 ----- 314
QY 5219 CTATTTTCCATAGTATTCAAGTTTGACAATGATCAAGTATTACTCTTTCTTTTCTTTT 5278
Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 315 ***PhePhe-----TyrPheLeu 320
QY 5279 TTTTCTTTTCTTTTGTGAGATGGAGTTTGGTCTTGTGCCCCATGTGCGAGTGAATGGA 5338
Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 321 PheLeuPhePheSerGluThrGluSerArgSer-ValAlaLysAlaGlyValGlnTrpAr 340
QY 5339 TCAYCATAGCTCAGCTGCAACCTCCACCTCCCTGGGTTCAGCAAAAGCTGTGCGCTCAGCT 5398
Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 340 gasPLeuGlySerLeuAlaProProGlyIlePheThrProPheSerCysLeuSerLeu 360
QY 5399 CCCGGGTAGATGGGATTACAGCGCCACACACACACCTCGGCTAATGTT-----TGTAT 5452
Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 360 uProSerSerTrpAspTyrArgArg-----ProHisTyrAlaArgLeuIlePheCysAl 378
QY 5453 TTTTAGTAGATGGGTTTCCACATGTTGGCCAGGCTGGTCTCAAACTCCCTGACCTC 5510
Db || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 378 ePheSerArgAspGlyValSerProTrp***ProGlyTrpSerArgSerProAspLeu 397
RESULT 7
E40201
artifact-warning sequence (translated ALU class E) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: E40201
R:Claverie, J.M.
personal communication, 1992
```


A:Reference number: A40201

A: Accession: E40201

A: Molecule type: DNA

A; Molecule type: DNA

A;Residues: 1-597 <CLA>

R;Claverie, J.M.

Genomics 12, 838-841, 1992

A;Title: Identifying coding exons by similarity search:

A; Reference number: A40

A:reference number: A40200; MUID:52241031; PMID:1372001
A:Contents: annotation

A;colcents: annotation
C:Comment: This "warning" entry is a conventional translation for all 6 words; a function of

C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of

Alignment Scores:

Pred. No.:	4,49e-25	Length:	597
Score:	325.00	Matches:	198
Percent Similarity:	33.12%	Conservative:	56
Best Local Similarity:	25.81%	Mismatches:	211
Query Match:	3.38%	Indels:	303
DB:	4	Gaps:	30

US-09-966-880A-9 (1-5514) x E40201 (1-597)

Qy	5508	GGTCAGGAGTTTGAGACCGCCTGGCCAAACATGGTGAAGACCCCATCTCTACTAAAAATAC	5449
Db	24	GlyTrpLupheGluThrSerLeuThrAsnMetGluLysProHisLeuTyr**LysTyr	43
Qy	5448	AAACATTAGCCGAGTGTGGTGGGGCCCTGTAATCCCATCTACCCGGGAGGCTGAG--	5391
Db	44	Lys-IleSerTrpAlaTrpTrpArgMetProValIleProAlaThrArgGluAlaGluTh	63
Qy	5390	-GCACACGCTTGGCTTCAACCAGGAGGTGGAGGTGTCAGTGAGCTATGRTCCGCACT	5332
Db	63	rGly-GluAsnCysLeuAsnProGlySerLysValCysGlyGlu--Ile--ValProSe	81
Qy	5331	CCACTCCAGCATGGCCAAACAGACCAAACTCCCATCTCAAAAAAATAAAAAA	5272
Db	81	rTyrSerGlyLeuGlyAsnLysSerLysThrProSerGln-LysLysLys*****A	101
Qy	5271	AAAGAAAGAGTAATACTTGTATCATCTGTCAAACTGGAATCTATGGAATAATAGAGACT	5212
Db	101	laArg-----	102
Qy	5211	CTAAATTTTAAATATTAAAGATTTTTTAAATCTCAAATATCATAGTGTCTAATAATAAT	5152
Db	102	-----	102
Qy	5151	AATTAGACTTAGTGTCPTAATGACAATGACTGTTTAAGATTCTTGTGTT--	5104
Db	103	-----CysGlySerSerCysLeu**SerGlnArgp	113
Qy	5103	ATTCTCTAGTCTTCTACAGGCACATATTACATGATTTGGATTATACATGATTCATGT	5044
Db	113	heGlySerSerArgArgGlyGly	120
Qy	5043	GTGATTTTGTACTCTGCTGCTTTTTTAACTAGTCAAAATGATCTAAATATCTTCCTATGT	4984
Db	121	-----SerProGluValGlySerL	127
Qy	4983	TGTTCCACATAATAGCTAGTATTTCTCCTCCTCTTTGTTTCTGGCTCCTAGCACAGTGC	4924
Db	127	euArgProAla**Pro-	132
Qy	4923	TTGCTATCATACCACTAATTTGAATAGCATTTGTGGAATAACTCTGGTAACCAAGTT	4864
Db	133	-----ThrTrpArg-----AsnProIleS	139
Qy	4863	TCCTTAAGAAATCAAG-	4834
Db	139	erThrLysAsnIleLysLeuAlaGlyArgGlyGlyAlaCysLeu**SerGlnLeuLeuG	159
Qy	4833	GCTGCACAGCACACACAAAGCTGTACAGCGTGGTTGATATGAAGACACAAGTATGCCTG	4774

```

Db      471 -----
QY      3951 GTGACAGACCCCTGACTTTAAAGTTAAATACCAAAATATACAGCTTAACCTAAGTAGCTCAT 3892
Db      472 -----ThrProAsnLeuArg***SerAlaProPro***ThrProLys-----Ala 486
QY      3891 CTAGGACAAACGCTCTTTGTAGTTGTTCCATTTATTTTAAATTTTCTTCTTATTTT 3832
Db      487 LeuGlyLeuGlnAla***AlaThrThrProGly*****PhePhePheLeu----- 504
QY      3831 TGAGGCAAGGTCTGGTCTGTCCAGGCTGGAGTGCGAGTGGCGCAAGCA----- 3781
Db      505 -----ArgArgSerPheAlaLeuValAlaGlnAlaGlyValArg 517
QY      3780 -----CTGCTCAGTCGACGCTTGACCTCTCGGCTCCAAAGATTCTC---CTGCTCAG 3730
Db      518 TrpHisAsnLeuThrAlaAsnPheAlaSerTrpValGlnAlaLeuSerCysLeuSer 537
QY      3729 CCCCTAGTAGTGGGACACGACATGCCACCATGCCGCTAATTTTGT----- 3675
Db      538 LeuProSerSerTrpAspTyrArgHisAlaProProArgProAlaAsn-PheilePheLe 557
QY      3674 -ATAGAAACAGGGTTTGGCATTGCTGCTCAGGCTGTGTGAAGTCCAGGGCTCAAGCGA 3616
Db      557 uValGluMetGlyPheLeuHisValGlyGlnAlaGlyLeuLysLeuProThrSerGlyAs 577
QY      3615 TCACCCAC-CTCAGCCTCCAGAGTGGGATTACAGCGTGAGCCACACACCTGGC 3557
Db      577 ProProArgLeuGluLeuProLysArgTrpAspTyrArgHisGluLeuProHisLeuAl 597
QY      3556 T 3556
Db      597 a 597

RESULT 8
A:0201
artifact-warning sequence (translated ALU class A) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: A40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: A40201
A:Molecule type: DNA
A:Residues: 1-627 <LA>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
A:Reference number: A40200; MUID:92241891; PMID:1572661
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of a
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of the
Alignment Scores:
Pred. No.: 6 52e-25 Length: 627
Score: 323.50 Matches: 172
Percent Similarity: 31.50% Conservative: 57
Best Local Similarity: 23.66% Mismatches: 191
Query Match: 3.36% Indels: 310
DB: 4 Gaps: 19

US-09-966-880A-9 (1-5514) x A40201 (1-627)
QY      5509 AGGTGAGAGTTGTGAGACGAGCTGCCCAACATGGTGAACCCCATCTCTACTAAAAATA 5450
Db      22 ArgSerGlyAspArgAspHisProGlyTyrHisGlyGluThrProSerLeuLeuLysIle 41
QY      5449 CAAACATTAGCCGAGTGTGGTGGCGCTGTATCCCATCTACCCGGAGCGTGGAG 5390
Db      42 GlnLysIleSerArgAla***CysGlyArgLeu***SerGlnLeuLeuGlyArgLeuArg 61

```

```

QY      5389 CGACAGCTTTGCTTGAACCCAGGAGCTGGAGTTGCAGTATGRTCATGTCATCC 5330
Db      62 GlnGluAsnGlyValAsnProGlyGlyAlaCysSerGluProArgSerArgHisCys 81
QY      5329 ACTCCAGCATGGCAACAAAGACCAAACTCCATCTCAAAAAAATAAAAAAATAA 5270
Db      82 ThrProAlaLeuAlaThr-GluArgAspSerValSerGluLysAsnLysAsnLys** 101
QY      5269 AGAAGAGTAATACCTTGATCATTTGCTCAAACTTGAATA-----CTATGGAAATAAG 5219
Db      101 *Lys*****AlaGlyArgGlyLeuThrProValIleProAlaLeuTrpGluAla 121
QY      5218 AAGAGCTCTAAATTTTAAATATTAAGATTTTTTAAATCTCAATATCATAGTGTCTAA 5159
Db      121 sAlaGly-----GlySe 125
QY      5158 TAATAATAATAGACTTAGTGTCTAATGACAATGACTGTTAAGATCTTGTGTTTATTC 5099
Db      125 rArgGlyGlnGluIleGluThrIleLeuAlaThrValLysProArgLeuTyr**Ly 145
QY      5098 TCTAGTCTTTCTACAGGCACATATTTACATGATTGATTATCTACTGATTCATGCTGTGAT 5039
Db      145 sTyrLysLysLeuAlaGlyArg----- 152
QY      5038 TTTGTACTCTGCTGCTTTTAACTAGTCAAAATGATCTAAATATCTTGTCTATCTGTTTC 4979
Db      153 -----SerAlaGly-----AlaCysSe 158
QY      4978 CCACATAATGACTGATTTCCTCCTCTTGTGTTTCTGCGTCTCAGACAGTGTCTTGCA 4919
Db      158 rProSerTyrLeuGlyGly**GlyArgArgMetAla**ThrArgGluAlaGluLeuAl 178
QY      4918 TATCAATAAGCACTAATTTGNAATAGCATTTG-----TGAATACTCTGGTACCCAGTTT 4863
Db      178 aValSerArgAspArgAlaThrAlaLeuGlnProTrpArgGlnSerGluThrProSer-- 197
QY      4862 CCTTAAAGAAATCAAGAGTCCCATTTGCAGCTCAGACACACAAAGCTGTACAGCGT 4803
Db      198 -ArgLysLysIleLysThrLysAsn----- 205
QY      4802 GGTGTATGAAGACACAGTGTATGCTGCTGATTTTTCCTCTAGTCTTTTCCAGAAATGCA 4743
Db      206 -----LysLys*****ProGly-----Al 213
QY      4742 TTTATCCAATTTCTGGAGCTAGTGTCTACCTTAATGGTACGGGAGAGTGCAGAGAA 4683
Db      213 aValSerArgLeu***Ser-----GlnHisPheGlyArgProArgArgAlaAspHI 230
QY      4682 T-----sGluValArgArgSerArgProSerTrpLeuProArg***AsnProValSerThrLysAs 250
Db      230 sGluValArgArgSerArgProSerTrpLeuProArg***AsnProValSerThrLysAs 250
QY      4667 GCTCAGGAATTTGAATAACGTTTGC----- 4643
Db      250 nThrLysAsn***ProGlyValValArgAlaProValValProAlaThrTrpGluAlaG 270
QY      4642 -----TGGCAAGAG----- 4634
Db      270 uAlaGlyGluTrpArgGluProGlyArgArgSerLeuGln**AlaGluIleAlaProLe 290
QY      4634 ----- 4634
Db      290 uHisSerSerLeuGlyAspArgAlaArgLeuArgLeuGlyLysLys**LysGlnLysIl 310
QY      4633 -----AATGAACTTTTCTTACACATTTATTTTCTTACATTGAGGTATATTTACA 4581
Db      310 eLysLys*****PhePheTyrPheLeuPheLeuPhePheSerGluThrGluSerAr 330
QY      4580 TACAGTAAACTCCACCCATTTTAAGATGCATAATTTCTACGAATATTTTGGATGTAAACAC 4521
Db      330 gSerValAlaLysAlaGlyValGln-----TrpArgAspLeuGlySe 344

```



```
Db 146 ---GlySerIleSerLeu***SerGlnLeu---LeuLysArgLeuGlyGlnGluAspHis 163
Qy 4278 -----
Db 164 LeuThrTrpGluValLysAlaGln**AlaMetIleAlaProLeuTyrSerSerLeuGly 183
Qy 4279 -----AGAGGCTCACACCTGTAATCC 4299
Db 184 AsnIleProArgProCysLeuLys*****ProSerMetAlaAlaArgLeu***Ser 203
Qy 4300 CAGCAGCTTTGAGAGCCAGCGTGGGAAGATTGCTAGAACTCAGGAGTTCAAGACAGCAGCTG 4359
Db 204 GlnHisPhe-ArgSerHisGlyArgCysMetLeuGlnLeuArgSerArgThrAlaTr 223
Qy 4360 GGCACACAGTGAACCTCATCTCCACAAAATGGGAAAAAGAAAGCAAAATCAGTGG 4419
Db 223 pAlaThr-CysArgAsnProThrSerThrLysLysArgTyrLysAsn***ProAspLeuV 243
Qy 4420 TTGTCCTGTGGGAGGGAAGGACTGCAAGAGGGAAGAAGCTCTGTGGGGTGGAGGTG 4479
Db 243 alAla***AlaCysSerProSerTyrSerLysGly-----TrpGly-Arg--- 257
Qy 4480 GTGATTGAGGTTCTGTATCCTGAGCTGTGTGAGCAGATTGGGGTGTTCATACCAAAAATA 4539
Db 258 -----ArgIleThr***Pro-----GlyLysSer---ArgLeuSerGluPro***Leu 272
Qy 4540 TTCGTAGAAATATGCAATCTTAAGTGGGTGAGTTTACTGATGTAATATTACCTCAATG 4599
Db 273 -----HisLeuCys---ThrProAlaTrpAla----- 280
Qy 4600 TAAGAAAAATAATGCTAGAAAAAGTTTCAATCTCTGCCAGCAACGTTATTCAAT 4659
Db 281 -----ThrTyrGlnAspProAla***Lys***** 289
Qy 4660 TCCTGAGCCCTTTACTTCGCAAAATCTCTGCACCTTCTGCCCTTACCATTTAGGTGACAG 4719
Db 290 -----*****PheSerGlyArgValLeuValCysCysPro-----GlyTrpSer 304
Qy 4720 ACTAGTCCCAAAATTTGGATA---AATGCAATTTCTGAAAGACTAGGGAC----- 4767
Db 305 ThrGluValGlnSerTrpLeuThrGluPro***LeuProArgSerSerAspProAla 324
Qy 4768 -----AAATCCAGGCATCACTTGTGCTTCATATCAACCA----- 4803
Db 325 ProAlaPhe***ValAlaGlyThrThrGlyLeuCysTyrGlnIleTrpLeuIlePheVal 344
Qy 4804 -----CGGTGTACAGCTTGTGTGTCTGTCTGAGCTGCAATGGGACTCTTGATTTCTT 4857
Db 345 SerPhePheCysArgGlyGlyValSerThrCys---CysProGlyCysSer----- 360
Qy 4858 TAAGGAACTTGGTTACCAGAGTATTTCCACAAATGCTATTCAAAATTAGTGTATTATGAT 4917
Db 361 -----*****ThrProGlu 364
Qy 4918 ATGCAAGACACTGTCTAGGAGCCAGAAACAAAGAGGAGAGAA---ATCAGTCAATTAT 4974
Db 365 LeuGluHisThrProAlaMetAlaProLysValLeuGlyLeuGlnAlaCysSerHis-Al 384
Qy 4975 GTGGGAACAACATAGCAAGATATTTAGATCATTTTGACTAGTTAAAAAAGCAGCAGAGTA 5034
Db 384 atrp*****Phe-----GlnAlaGlySerTr 394
Qy 5035 CAAAATCACACATGCAATCAGTATATCAAAATCATGTAATATGTCCTGTAGAAAGAC 5094
Db 394 pTyrValAlaGlnAla-GlyVal-----GlnArgCysAsnHis----- 406
Qy 5095 TAGAGGAATAAACACAAGAATCTTAACAGCTATTGTCAATTAGAC----- 5138
Db 407 -----GlySerLeuSerLeuAspPheProGlyGlnValI 418
Qy 5139 -----ACTAAGTCTCAATT 5151
Db 418 leuLeuProGlnProPheGlu***LeuGlyLeuGlnAlaTyrAlaThrArgSerGly* 438
```

```
Qy 5152 ATTATTATTAGACACACTATGATATTGAGATTTAAAAAATCTTTAATATTTTAAATTTAG 5211
Db 438 **Phe-----LeuT 441
Qy 5212 AGCTCTTCTATTTTTCATAGTATTCAAGTTTGACAATGATCAAGATATTACTCTTTCTTT 5271
Db 441 yrLeuPhePheValGluValGlyPheArgHisValAlaGlnAlaValLeuGluLeuLeu 461
Qy 5272 TTTTTTTTTTTTTTTTT----- 5288
Db 461 erTrpSerIleHisLeuProTrpLeuLeuLysCysTrpAspTyrArgArgAlaAlaMetL 481
Qy 5289 -----TTTTTGGATGGAGTTTGTGCTCTTGTTCGCATGCTGGAGTGAATGGC 5337
Db 481 euGly*****PheArgGlnGlyLeuGly-MetLeuProArgLeuGluTyrArgGly 500
Qy 5338 ATGAYCATAGCTCACTCAACCTCCACCTCCTGGTTTCAAGCAAGCTGTGCGCTCAGCC 5397
Db 501 AlaIleMetAlaHis***AlaLeu-ThrSerGlnValLys***SerSerCysPro-SerL 520
Qy 5398 TCCCGGCTAGATGGGATTACAGCGCCGCCACACACTCGGCTCTCAAACTCC 5503
Db 520 euLeuSerSerTrpAspTyrArgLeuMetLeuProAspLeuAlaAsnPheCysIlePheP 540
Qy 5458 GTAG-----AGATGGGGTTTCCACCATGTTGGCCAGCGCTGGTCTCAAACTCC 5503
Db 540 heLeu***ArgTrpGlyPheAspMetLeuProArgLeuPheLeuAsnSer 556
```

RESULT 10

S41044

Chromosomal protein - human

C:Species: Homo sapiens (man)

C:Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 21-Jan-2000

C:Accession: S41044

R:Yeo, J.P.; Alderuccio, F.; Toh, B.H.

Nature 367, 288-291, 1994

A:Title: A new chromosomal protein essential for mitotic spindle assembly.

A:Reference number: S41044; MUID:94166884; PMID:8121495

A:Accession: S41044

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-418 <YEO>

A:Cross-references: GB:I26953; NID:g537529; PIDN:AAB68050.1; PID:g537530

C:Superfamily: human 48.2K chromosomal protein

C:Keywords: chromosomal protein

Alignment Scores:

Pred. No.:	1.97e-14	Length:	418
Score:	224.00	Matches:	91
Percent Similarity:	37.39%	Conservative:	32
Best Local Similarity:	27.66%	Mismatches:	74
Query Match:	2.32%	Indels:	132
DB:	2	Gaps:	17

US-09-966-880A-9 (1-5514) x S41044 (1-418)

```
Qy 3550 ACAGACAGCCAGGTG---TGG-----TGGCTCACGCCTGTAATCCACGACCTCGGAG 3600
Db 163 ThrGlnSerTyrValAspTrpGlyArgTrpLeu-----IleTrpGlu 176
Qy 3601 GCTGAGGTGGGTGATCGCTTGAGCCCTTGAGCCCTGAGACACGCTGAGCAACATGCGCA 3660
Db 177 AlaLysAlaGlyGluSerLeuGluValArgSerArgProAlaSerGlnSerArgArg 196
Qy 3661 AACCCCTGTTTCATA-----ACAAAAATT 3684
Db 197 AsnSerValSerThrLysAsnIleLysIleSerProValSerThrLysAsnIleLysIle 216
Qy 3685 AGCCGGCATGGTGG-----CATGTGCTGTGCTGCCAGCTACTAGGGGG 3729
Db 217 SerGlnThrTrpTyrLeuPheGlyGlyValHisLeu---LeuValProThrThrArgAsp 235
```



```
Db 263 yLysSerArgLeuSerGluPro***LeuHisLeuCysThrProAlaTrpAla----- 280
QY 4715 TCACCTAAGTGTACGGGGCAGAAAGTGCAGAGAATTTTCGGAAGTAAAGGGCTCAGGAATTT 4656
Db 281 -----ThrTyrGlnAspProAla***Lys*****PheSerGly---ArgValLe 297
QY 4655 GAATAACGTTTGC-----TGGCAAGAGAATTGAAACTTTTCTTACACATATATTTT 4605
Db 297 u-----ValCysCysProGlyTrp----- 303
QY 4604 TCTTACATTGAGGTATAATTTACATACAGTAAACTCCACCATTTCAGCATATTCCT 4545
Db 304 -----SerThrGluValGln----- 308
QY 4544 ACGAATATTTTGGATGTAACACCCCAAACTGCTACCACAGTCAGGATACAGAACTGA 4485
Db 309 -----SerTripleLeuThrGluPro** 315
QY 4484 ATCACCACTCTACCCACCACAGAGCTTCTCCCTCTTTGAGTCCTCCCTCCCTCCACAG 4425
Db 315 *LeuProArgSerSerAsp-----ProProAlaProAlaPhe** 328
QY 4424 GACA-----ACCACGTATTG-----CTTCTCTTTTTCCTCCATTTTGT 4386
Db 328 *ValAlaGlyThrThrGlyLeuCysTyrGlnIleTrpLeuIlePheValSerPhePheC 348
QY 4385 GGAGAATGAGTTTCACTGTGTGGCCAGCGTGTCTGAACCTCCAGCTTCTAGCAATC 4326
Db 348 sArgGlyGlyValSerThrCysCysProGlyCysSer***ThrProGluLeu----- 365
QY 4325 TTCCCAAGCGCTCTCAAGAGCTGGGATTACAGGTGTGAGCCCTATGCGCTGCCTGA 4266
Db 366 -GluHisThrProAlaMetAlaProLysValLeuGly-----LeuGlnAlaCys----- 381
QY 4265 TTTGCTTTCTGTAACATAAGTTACTTTTCATTTTCTAGAAATTTGCTATGCTATACATAT 4206
Db 382 -----SerHisAlaTrp*****PheGlnAlaGly----- 392
QY 4205 CTCCTCTTCATGCTGACTCTTTTCATTTAGCGTCAGCATTTGGG----- 4160
Db 393 -SerTrpTyrValAlaGlnAlaGlyValGlnArgCysAsnHisGlySerLeuSerLeuAs 412
QY 4159 -----ATTCATCCCAAGTAGCTTCTCTCTTGTACTCCAGCGGTGG 4119
Db 412 pPheProGlyGlnValIleLeuLeuProGlnProPheGlu***LeuGlyLeuGlnAlaTy 432
QY 4118 TATTCATTCGATGATATAGACATTTTGTATTATTTGTCTACTTTGTTGATGGATATTG 4059
Db 432 rAlaThrArgSerGly***PheLeuTyrLeuPhePheValGlu----- 446
QY 4058 GATAGTCCTACTTTGGGGTTTGTGAATAAAGGTACTATGAAACATTCACGTACAGTCT 3999
Db 447 -----ValGlyPhe-----ArgHisValAlaGlnAlaValLeuGluLe 459
QY 3998 TTGTGTGGATATACATTGAGAATTGATATTTTAATAATCTAGTGCAGGTGACAGACCTG 3939
Db 459 u----- 459
QY 3938 ACTTTAAAGTTAAATACAAAATATCAGCCTAACTAACTAGCTCATCTAGGACAACAGC 3879
Db 460 -----LeuSerTrpSerIleHisLeuProTr 468
QY 3878 TCTTTGTAGTCTTCTCCATTATTTATTTTCTCTCTT-----ATTTTTCAGGCA 3825
Db 468 pleuLeuLysCysTrpAspTyrArgArgAlaAlaMetLeuGly*****PheArgAl 488
QY 3824 AGGTCTGTCTCTCACCAGCGCTCGAGTGCAGTGCAGCAAGCACTGCTCAGTCAGCGCT 3765
Db 488 nGlyLeuGlyMetLeuProArgLeuGluTyrArgGlyAlaIleMetAlaHis-***AlaAl 508
QY 3764 TGACCTCTCGGCTCCAAAGATTCCTCTGCCTCAGCCCTAGTAGCTGGGACCAAGGC 3705
Db 508 euthrSerGlnValLys***SerSerCysProSerLeuLeuSerSerTrpAspTyrArgL 528
```

```
QY 3704 ACATGCCCATCGCCGGCTAAATTTTGTATTATAG-----AAACAGGTTTCCCA 3655
Db 528 euMetLeuProAspLeuAlaAsnPheCysIlePhePheLeu***ArgTrpGlyPheAspM 548
QY 3654 TGTGTCTCAGCGCTGGTCTTGAACCTCCAGGCTCAAGCGATCCACCCACCTCAGCCTCCCA 3595
Db 548 etLeuProArgLeuPheLeuAsnSer***AlaGlyAlaTyr--ThrCysHisGlySer** 567
QY 3594 GAGTCTGGGATTACAGCGCTG 3573
Db 567 *SerAlaGlyIleThrGlyVal 574

RESULT 12
A46010
X-linked retinopathy protein (C-terminal, clone XEH.8c) - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A46010
R:Wong, P.; MacDonald, I.M.; Sood, R.; Smith, C.; Pilon, R.; Tenniswood, M.
Genomics 15, 467-471, 1993
A:Title: Identification and partial characterization of a candidate gene for X-linked
A:Reference number: A46010; MUID:93224131; PMID:8468040
A:Accession: A46010
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-100 <WON>
A:Cross-references: GB:S58722; NID:q299470; PIDN:AAB26149,1; PID:q299471
A:Note: sequence extracted from NCBI backbone (NCBIN:129339, NCBIP:129340)

Alignment Scores:
Pred. No.: 1,94e-13 Length: 100
Score: 213.50 Matches: 49
Percent Similarity: 71.95% Conservative: 10
Best Local Similarity: 59.76% Mismatches: 20
Query Match: 2.21% Indels: 3
DB: Gaps: 1

US-09-966-880A-9 (1-5514) x A46010 (1-100)

QY 5276 TTTTTTTTTTTTTTTTTTTCAGATGGAGTTTGTGCTTGTGCCATGCTGAGTGGCAATG 5335
Db 1 PhePhePhePhePhePheGluThrGluSerCysSer-ValAlaGluAlaGlyValGlnTr 20
QY 5336 GCATGAYCATAGTACTGCAACCTCCACTCTCTGGTTTCAAGCAAGCTGCGCTCAG 5395
Db 20 pCysAspLeuGlySerLeuLysSerProProGlySerSerAspSerProAlaSerA 40
QY 5396 CCTCCCGGTAGATGGGATTACAGCGCCCAACACACTCGGCTAATGTTGTATTTT 5455
Db 40 laSerArgValAlaGlyIleThrGlyMethHisHisThrGlnLeuIlePheValPheL 60
QY 5456 TAGTAGAGATGGGTTTTCAC---CATGTTGGCCAGCGCTGCTCAAACCTCTGACCTCAG 5512
Db 60 euValGluThrGlySerHisMetGlnLeuSerAspSerThrLeuValIleThrThrAlaG 80
QY 5513 AG 5514
Db 80 ln 80

RESULT 13
S41044
chromosomal protein - human
C:Species: Homo sapiens (man)
C:Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 21-Jan-2000
C:Accession: S41044
R:Yeo, J.P.; Alderuccio, F.; Toh, B.H.
Nature 367, 288-291, 1994
A:Title: A new chromosomal protein essential for mitotic spindle assembly.
A:Reference number: S41044; MUID:94166884; PMID:8121495
A:Accession: S41044
A:Status: preliminary
A:Molecule type: mRNA
```

Db	2	ArgAlaTrpTrpLeuThrSerValIleProAlaLeuTrpAspAlaGluValGlyGlyLeu	21
Qy	3619	CTTGAGCCCTGGAGTTCACAGCAGCAGCTGAGCAACATGGCAAAACCTGTTCTATAACA	3678
Db	22	LeuGluThrArgSerArgProAlaTrpAlaThr***GlnAspProSerSerIle-Ly	41
Qy	3679	A-----AAATTAGCCGGGCATGGTGGCATGTGCTGTGGTGC	3714
Db	41	silLeLys***Asn**LeuGly*****AlaGlyHisGlyGly***HisLeu***Se	61
Qy	3715	CCAGCTACTAGGGGCTGAGCAGGAGCAATCTTTTGGAGCCCGAGGAGTGCAGGCTGCAC	3774
Db	61	rgIn-----HisPheGlyMetLeuLys***Glu-----	70
Qy	3775	GAGCAGTGTCTGGCCCACTGCATCCAGCTCGGTGACAGGACCAAGACCTTGCTCAAAA	3834
Db	71	-----AspCysLeuArgProGlyValGlnAspGln-----	80
Qy	3835	AAATAAGAGAAAAATAAATAATATGGAACAACACTACAAAGAGTGTGTCCTAGATG	3894
Db	80	-----	80
Qy	3895	AGCTACTTAGTGTAGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGTCACCTG	3954
Db	81	-----LeuGly-----	82
Qy	3955	CACTACATATTAAATAATCAATTCTCAATGTATATCCACACAAGACTGGTACGTGAAT	4014
Db	83	-----GlnHisSerLysThr-----	87
Qy	4015	GTTCATAGTACCTTTATTCAAAAACCCCAAGTAGAGACTATCCAAATATCCATCAACA	4074
Db	88	-----ProAlaLeu***Lys-----	92
Qy	4075	AGTGAACAATAACAAATATGCTATATCCATGCAATGCAATACCACCCTGCAGTACAA	4134
Db	93	***AsnLys-----IleSerTrpVal*****Gln-----	103
Qy	4135	AGGAAGAAGTACTTGGGGATGAATCCCAAAGTCATGACGCTAAATGAAGAGTGCAGACA	4194
Db	103	-----	103
Qy	4195	TGAAGGAGGAGATAATGTATGCCATACGAAATCTTAGAAAAATAAGTAATATAGTTA	4254
Db	103	-----	103
Qy	4255	CAGAAACCAATCAGGCGAGGCATAGAGGCTCACACTGTAAATCCAGCACTTTGAGAGG	4314
Db	104	-----GlyMetValAlaAsnIleCysAsnProSerThrLeuGlyCy	117
Qy	4315	-CCACGTGGGAAGTGTCTAGAAGTCTCAGGAGTTCACAGACCGCTGGGGCAACACAGTAA	4373
Db	117	***SerArgArgIleAla***AspGlnGluPheLysThrSerLeuGlyAsnIleAlaAr	137
Qy	4374	ACTCCATCTCCACAAAAATGGGAAAAAGAAAGCAAAATCAGTGGTGTGCTGTGGGA	4433
Db	137	g-----ProGlnLeuTrpLysAsnLys-----	144
Qy	4434	GGGGAAGGACTGCAAGAGGGAAGAACCTCTGTGGGTGAGGTGGTGTATTCAGGTTCT	4493
Db	145	-----IleLysLeu-----	147
Qy	4494	GTATCCTGACTGTGGTAGCAGTTTGGGGTGTTTACATCCAAAAATATTCGTAGAAATTATG	4553
Db	148	-----AlGly*****HisProIleAsnPhelle-----Le	159
Qy	4554	CATCTTAAATGGGTGGAGTTTACTGTATGTAAATATATACCTCAATGAAGAAAAATAAT	4613
Db	159	uPheLeu***SerTrpGlyLeu-----	166
Qy	4614	GTGTAGAAAGTTTCAATCTCTGCGCAGCAACAGTTATTCAAATTCCTGAGCCCTTTA	4673
Db	167	-----AlaMetLeuProLysLeuValLeuAsnSer-----	176

```

A:Molecule type: DNA
A:Residues: 1-196 <RES>
A:Cross-references: EMBL:X55777; NID:g288143; PIDN:CAA39297.1; PID:g288145

Alignment Scores:
Pred. No.:      4,11e-11
Score:          192.00
Percent Similarity: 65.88%
Best Local Similarity: 54.12%
Query Match:      1.99%
DB:              2

US-09-966-880A-9 (1-5514) x 138022 (1-196)

Qy      5512 CTGAGTCCAGGAGTTTGAGACCAGCCTGCCAACATGGTGAACCCCATCTCTACTAAAA 5453
Db      4   LeuArgSerGlyValGlnAsnProGlyGlnHisGlyLysIleProSerLeuLeuLys 23

Qy      5452 ATACAAACATTAGCCGAGTCTGGTGTGGCGCCTGTAATCCCATCTACCCGGAGGCGTG 5393
Db      24   IleGlnGluLeuAlaGlyHisGlyGlyArgCysLeuGlnSerGlnLeuLeuArgArgLeu 43

Qy      5392 AGGGCAGCATTGCTTGTAACCCAGGAGGTGGAGGTTGCAGTGCAGCTATGTCATGCCAT 5333
Db      44   ArgGlnGlnAsnHisLeuAsnSerGlyGlyArgGlyCysSerGluProLysSerHisLeu 63

Qy      5332 TCCACTCCAGCATGGGCAACAGACCAAACTCCATCTCAAAAAAATAAAAAAATAAAAA 5273
Db      64   CysIleProAlaTrpValThr-GluGlyAspSerValSerLysGlnAsnLysThrLysAs 83

Qy      5272 AAAAGAAGAGATA 5260
Db      83   nGluGlnHisLeu 87

```

Search completed: June 14, 2003, 18:57:31
Job time : 214.707 secs

4674 CTTGCGAAATCTCTGCACTTCTGCCCCGTACCATTTAGGTGACAGCACTAGCTCCACAA 4733

176 ----- 176

4734 TTGGTAATATGCATTTCTGGAAAAGACTAGGGACAAAATCCAGGCATCACTTGTGCTTTC 4793

177 -TrpSerGlnAlaIleLeuLeuGlnHisProLysVal- 189

4794 ATATCAACCACGCTGTACAGCTTGTGCTGTCTGCTGTGCTGCAAGTGCATGGGACTCTTGATT 4853

189 ----- 189

4854 TCTTTAAGGAACTTTGGTTTACCAGAGTATTTCCACAAATGCTATTCAAATTAGTGCTTA 4913

190 -----LeuGlyLeuGlnMetLeuAlaThr- 197

4914 TGATATGAAGACACTGTGCTAGGAGCCAGAAAACAAGAGGAGAGAAATCAGTCATTA 4973

197 ----- 197

4974 TGTGGGAACAACATAGCAAGATATTTAGATCATTTTGACTAGTTTAAAAAAGCAGCAGAGT 5033

197 ----- 197

5034 ACAAATCACACATGCAATCAGTATATATCCAAATCATGTAATATGTGCCTGTAGAAAGA 5093

197 ----- 197

5094 CTAGAGGAATAAACACACAAGAACTCTTAACAGTCATTTGTTCATTAGACACTTAAGTCTAATTAT 5153

198 -----MetProCys*****ThrGlnLeuIleLeu 208

5154 TATTATTAGACACTATGATATTTGAGATTTAAAAAATCTTTAAATTTTAAAAATTTAGAG 5213

208 u----- 208

5214 CTCCTCTATTTTCCATAGTATTCAAAGTTTGACAATGATCAAGTATTACTCTTCTTTT 5273

208 ----- 208

5274 TTTTNTTTTTTTTTTTTTTTTGTGATGGAGTTTTGGTCTGTGTGCCCATGCTGGAGTGGAA 5333

209 -----PheTyrPheTyrArgAlaGlyValLeuLeu-CysCysProSerTrpSer- 224

5334 TGGCATGATCAGCTCACTGCTCAACCTCCACCTCGGTTCACGCAAGCTGTCGCCTC 5393

225 -----***Thr-ProGlyLeuLysGlnSerSerTyrPhe 235

5394 AGCCTCCCGGTAGATGGATTACAGCGCCACCA-----CAC 5435

236 SerIleProLysCysTrpAspTyrArgCys***ProProCysProAla*****Pro 255

5436 TCGGCTAATGTTGTATTTTATTAGTAGATGGGGTTTCACCATGTTGCCAGCTGGTCT 5495

256 Ser***PheTyrPheIlePhe-IleGluLeuGlySerCysTyrValAlaGlnAlaGlyLe 275

5496 CAAACTCTCTGACCTCA 5511

275 uGluLeuValSer 280

RESULT 15

I38022

hypothetical protein - human

C:Species: Homo sapiens (man)

C>Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 05-Nov-1999

C:Accession: I38022

R:Yang, S.S.; Zhang, K.; Vieira, W.; Taub, J.V.; Zellstra-Ryalls, J.H.; Somerville, R.L.

Cancer, Res. 50, 5658-5667, 1990

A:Title: A human hepatocellular carcinoma 3.0-kilobase DNA sequence transforms both rat

A:Reference number: I38021

A:Accession: I38022

A>Status: preliminary; translated from GB/EMBL/DBJ

RESULT 15
I38022
hypothetical protein - human
C:Species: Homo sapiens (man)
C:Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 05-Nov-1999
C:Accession: I38022
R:Yang, S.S.; Zhang, K.; Vieira, W.; Taub, J.V.; Zellstra-Ryalls, J.H.; Somerville, R.L.
Cancer Res. 50, 5658-5667, 1990
A:Title: A human hepatocellular carcinoma 3.0-kilobase DNA sequence transforms both rat
A:Reference number: I38021
A:Accession: I38022
A>Status: preliminary; translated from GB/EMBL/DBJ

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:11:13 ; Search time 69.1643 Seconds
(without alignments)
6613.252 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 9659
Sequence: 1 acagacgaatacatagtccca.....tcaactctcgacctcagag 5514

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cgcn2_1/USPRO_pool/US09966880/runat_14062003_175524_10304/app_query.fasta_1.9493
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09966880_ECGN_1_196_brnat_14062003_175524_10304 -NCPU=6 -ICPU=3
-NO_MMAPP -LARGESQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	554	5.7	593	1 ALU7_HUMAN	P39194 homo sapien
2	553	5.7	591	1 ALU1_HUMAN	P39188 homo sapien
3	526	5.4	593	1 ALU6_HUMAN	P39193 homo sapien
4	522.5	5.4	591	1 ALU8_HUMAN	P39195 homo sapien
C 5	490.5	5.1	591	1 ALU1_HUMAN	P39188 homo sapien
C 6	489.5	5.1	593	1 ALU7_HUMAN	P39194 homo sapien
C 7	458.5	4.8	591	1 ALU8_HUMAN	P39195 homo sapien
C 8	448	4.7	593	1 ALU6_HUMAN	P39193 homo sapien
9	431.5	4.5	585	1 ALU5_HUMAN	P39192 homo sapien
10	426.5	4.4	587	1 ALU2_HUMAN	P39189 homo sapien
11	410.5	4.3	587	1 ALU2_HUMAN	P39189 homo sapien
12	385	4.0	587	1 ALU3_HUMAN	P39190 homo sapien
C 13	376.5	3.9	585	1 ALU5_HUMAN	P39192 homo sapien
C 14	362.5	3.8	587	1 ALU3_HUMAN	P39190 homo sapien
C 15	310.5	3.2	603	1 ALU4_HUMAN	P39191 homo sapien
C 16	255.5	2.7	603	1 ALU4_HUMAN	P39191 homo sapien
17	224	2.3	418	1 YY1_HUMAN	P49646 homo sapien
C 18	204.5	2.1	418	1 YY1_HUMAN	P49646 homo sapien

19	139.5	1.4	124	1	YY3_HUMAN	P20931 homo sapien
C 20	136	1.4	841	1	NEK4_HUMAN	P51957 homo sapien
C 21	129.5	1.3	629	1	2195_HUMAN	O14628 homo sapien
C 22	127	1.3	841	1	NEK4_HUMAN	P51957 homo sapien
C 23	122.5	1.3	881	1	PRP2_HUMAN	Q99959 homo sapien
C 24	113	1.2	619	1	REL_HUMAN	Q04864 homo sapien
C 25	109.5	1.1	501	1	GYG2_MOUSE	O15488 homo sapien
C 26	107	1.1	465	1	SEP3_MOUSE	Q921S5 mus musculu
C 27	105	1.1	1371	1	UBPJ_HUMAN	O94966 homo sapien
C 28	99	1.0	77	1	Q300_MOUSE	Q02722 mus musculu
C 29	99	1.0	369	1	TAZR_HUMAN	P21731 homo sapien
C 30	98	1.0	438	1	SGCE_HUMAN	O43556 homo sapien
C 31	98	1.0	447	1	KBF3_HUMAN	Q04860 homo sapien
C 32	97.5	1.0	590	1	NU5M_TRYBB	P04540 trypanosoma
C 33	96	1.0	153	1	YH17_YEAST	P38898 saccharomyc
C 34	95.5	1.0	1259	1	EH01_HUMAN	P48553 homo sapien
C 35	93	1.0	169	1	YJ85_YEAST	P47152 saccharomyc
C 36	91.5	0.9	491	1	YEAL_ECOLI	P76236 escherichia
C 37	91.5	1.0	533	1	YADC_SCHPO	Q09837 schizosacch
C 38	90	0.9	447	1	KBF3_HUMAN	Q04860 homo sapien
C 39	90	0.9	718	1	RHG8_HUMAN	Q9nsq0 homo sapien
C 40	88	0.9	453	1	VG_DROME	Q26366 drosophila
C 41	88	0.9	741	1	RED1_HUMAN	P78563 homo sapien
C 42	87.5	0.9	319	1	TCB1_RABIT	P06333 oryctolagus
C 43	87	0.9	510	1	YH43_YEAST	Q03218 saccharomyc
C 44	86	0.9	230	1	UREE_YEREN	P42869 versinia en
C 45	86	0.9	494	1	HMBC_DROME	P09081 drosophila

ALIGNMENTS

RESULT 1					
ALU7_HUMAN					
ID	ALU7_HUMAN	STANDARD;	PRT;	593 AA.	
AC	P39194;				
DT	01-FEB-1995 (Rel. 31, Created)				
DT	01-FEB-1995 (Rel. 31, Last sequence update)				
DE	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Alu subfamily SQ sequence contamination warning entry.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=95021756; PubMed=7935834;				
RA	Claverie J.-M., Makalowski W.;				
RT	"Alu alert."				
RL	Nature 371:752-752(1994).				
RN	[2]				
RP	CONCEPT.				
RX	MEDLINE=92241891; PubMed=1572661;				
RA	Claverie J.-M.;				
RT	"Identifying coding exons by similarity search: alu-derived and other potentially misleading protein sequences.";				
RL	Genomics 12:838-841(1992).				
RN	[3]				
RP	ALU FAMILIES CLASSIFICATION.				
RX	MEDLINE=88333009; PubMed=3138422;				
RA	Quentin Y.;				
RT	"The Alu family developed through successive waves of fixation closely connected with primate lineage history.";				
RL	J. Mol. Evol. 27:194-202(1988).				
RN	[4]				
RP	ALU FAMILIES CLASSIFICATION.				
RX	MEDLINE=91178815; PubMed=1706781;				
RA	Jurka J., Milosavljevic A.;				
RT	"Reconstruction and analysis of human Alu genes.";				
RL	J. Mol. Evol. 32:103-121(1991).				
CC	-!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU				

CC REPEATS.
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14573; -; NOT_ANNOTATED_CDS.
CC KW Hypothetical protein.
CC FT DOMAIN 1 97 FRAME-1.
CC FT DOMAIN 101 196 FRAME-2.
CC FT DOMAIN 200 295 FRAME-3.
CC FT DOMAIN 299 395 FRAME-4.
CC FT DOMAIN 399 494 FRAME-5.
CC FT DOMAIN 498 593 FRAME-6.
CC SQ SEQUENCE 593 AA; 64417 MW; 54A4F50F33A6089F CRC64;

Alignment Scores:
Pred. No.: 2,33e-48 Length: 593
Score: 554.00 Matches: 194
Percent Similarity: 33.13% Conservative: 28
Best Local Similarity: 28.96% Mismatches: 130
Query Match: 5.74% Indels: 320
DB: 1 Gaps: 16

US-09-966-880A-9 (1-5514) x ALU7_HUMAN (1-593)
QY 3568 TGGCTCACGCTGTAATCCACGACTCTGGGAGGCTGAGTGGTGGATCGTGTGAGCC 3627
Db 5 TrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlySerProGluVal 24
QY 3628 TGGAGTTCAGACCGCTGAGCAACATGGCAAAACCCCTGTCTCTATA-----ACAAAA 3681
Db 25 ArgSerArgProAlaTrpProThrTrp***AsnProValSerThrLysAsnThrLys 44
QY 3682 ATTAGCGGGCATGTGGCATGTGCTGTGGTCCGACGCTACCTAGGGG-GCTCAGGCAGGA 3740
Db 45 IleSerArgAlaTrpTrpArgAlaProValIleProAlaThrArgGluAlaGluAlaGly 64
QY 3741 GAATCTTGGAGCCAGGAGGTCAGGCTGCACTGAGCAGTGTCTCCGCGCACTGCCTCC 3800
Db 65 GluSerLeuGluProGlyArgArgArgLeuGln***AlaGluIleAlaProLeuHisSer 84
QY 3801 AGCCTGGGTGACAGGACACGACTTGCCTCAAAAAAATAAGAGAAAAATTAATAATAA 3860
Db 85 SerLeuGlyAsnLysSer-GluThrProSerGlnLys----- 96

QY 3861 TCGAAACAACACTACAAAGAGCTGTTGTCCTAGATGAGCTACTTAGTTAGGCTGATATTTTG 3920
Db 96 ----- 96
QY 3921 GTATTTAACTTTTAAAGTCAGGGTCTGTCCACCTGCACCTACATATTATAAATATCAATTCT 3980
Db 96 ----- 96
QY 3981 CAATGTATATCCACACAAAGACTGGTACCTGAATGTTTCATAGTACCTTTTATTACAAAAAC 4040
Db 96 ----- 96
QY 4041 CCCAAAGTAGAGACTATCCAAATATCCATCAACAAGTGAACAATAAACAATAATGTCGTA 4100
Db 96 ----- 96
QY 4101 TATCCATGCAATGGAATACACCCCTGCAGTACAAAGGAAGAAAGCTACTTTGGGGATCAATC 4160
Db 96 ----- 96
QY 4161 CCAAAGTCATGACGCTAAATGAAAGAGTCAGACATGAAGGAGGAGATAATGTATGCCATA 4220
Db 96 ----- 96
QY 4221 CGAAATTCAGAAATGAAAGTAACATTATGTACAGAAAGCAAAATCAGGGCAGGCATAG 4280
Db 97 -----Lys*****AlaGlyArgG1 104
QY 4281 AGGCTCACACCTGTAAATCCACGACTTTGAGAGGCCACCTGGG-GAAGATTGCTAGAACATC 4339
Db 104 yGlySerArgLeu***SerGlnHisPheGlyArgProArgValAspHisLeuArgSe 124
QY 4340 AGGAGTTCAGACCCAGCTGGGCAACACAGTGAACCTCCATTCACCAAAAATGGGAAA 4399
Db 124 rGlyValArgAspGlnProGlyGlnHisGlyGluThrPro-SerLeuLeuLysIleGlnL 144
QY 4400 AAAAGAAACAAATCAAGTCAGTGGTGTCTGTGGGGAGGGAAGGACTCCAAGAGGGAAGAA 4459
Db 144 ys-----LeuAlaGlyArg- 148
QY 4460 GCTCTGGTGGGTGAGGTGGTTCATTCAGTGT-----C 4492
Db 149 -----GlyGlyArgLeu***SerGlnLeuLeuLeuGlyArgLeuArgGlnGluAsnArgL 167
QY 4493 TGTATCTCCTGCTGTGGTAGCAGTTTGGGGTGTGTACATCCAAAAATATTCGTAGAAATAT 4552
Db 167 euAsnProGlyGlyGly-----GlyCysSerGluProArg-----SerArgH 181
QY 4553 GCATCTTAATGGTGGAGCTTTACTGTATGTAAATATACCTCAATGTGAAGAAAAATAA 4612
Db 181 isCysThrProAlaTrpAlaThrArg-----AlaLysLeuA 193
QY 4613 TGTGTAAGAAAAGTTTCAATTCCTTTGCCAGCAACGTTTATTCAAAATTCCTGAGCCCTTT 4672
Db 193 rgLeuLysLys*****ProGlyAla----- 202
QY 4673 ACTTCGCAAAATTCCTGCTGCACCTTCGTCG-----CCCGTACCATTTAGGTGACAGCACTAGCTCCA 4729
Db 203 -----ValAlaHisAlaCysAsnProSerThrLeuGly-----GlyArgG 216
QY 4730 CAATTTGGATAAATGCATTTCTGGAAAAGACTAGGGCAAAATCCAGGCATCCTGCTGTCG 4789
Db 216 lyGlyTrpIleThr***GlyGlnGlu-----PheGluThrSerLeuAlaA 231
QY 4790 TTTTCATATCAACACAGCTGT-----ACAGCTTGTGTGCTGTGCTGCA 4831
Db 231 snMetValLysProArgLeuTyr***LysTyrLysAsn***ProGlyValValAlaGlyA 251
QY 4832 GCTGCAATGGGAGCTCTTTCATTTTAAAGAAACTTGGGTACCAGAGATTTTCCACAA 4891
Db 251 laCysAsnProSer-----TyrSerGlyGly***GlyArgArgIleAla***ThrA 268
QY 4892 ATGCTATTCAAAATAGTGTCTTATGATATGCAAGACACTGTGCTAGGAGCCAGAAAACAA 4951

RT closely connected with primate lineage history." ;
RL J. Mol. Evol. 27:194-202(1988) .
RN [4]
RN ALU FAMILIES CLASSIFICATION.
RP MEDLINE-91178815; PubMed-1706781;
RX Jurka J., Milosavljevic A.;
RA "Reconstruction and analysis of human Alu genes." ;
RT J. Mol. Evol. 32:105-121(1991).
RL
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFICIALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and/or commercial
CC

CC entities requires a license agreement (see <http://www.rsp-sib.ch/announce/>) or send an email to license@sib.ch

```

CC      OI send an email to license@rsb.sis.chi.
-----
CC      EMBL; U14567; -; NOT_ANNOTATED_CDS.
KW      Hypothetical protein.
FT      DOMAIN 1 96 FRAME-1.
FT      DOMAIN 100 195 FRAME-2.
FT      DOMAIN 199 294 FRAME-3.
FT      DOMAIN 298 393 FRAME-4.
FT      DOMAIN 397 492 FRAME-5.
FT      DOMAIN 496 591 FRAME-6.
SQ      SEQUENCE 591 AA; 63790 MW; 665D395735519D95 CRC64;

Alignment Scores:
Pred. No.:      2,95e-48      Length:      591
Score:          553.00      Matches:    218
Percent Similarity: 37.81%      Mismatches: 55
Best Local Similarity: 30.19%      Gaps:       22
Query Match:    5.73%
DB:             1

US-09-966-880A-9 (1-5514) x ALU1_HUMAN (1-591)

QY      3568  TGGCTCAGCGCTGTATCCAGACACTCTGGGAGGCTGAGGTGGGTGGATCGCTTGAGCCC 3627
Db      5      TrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlySerLeuGluPro 24
QY      3628  TGGAGTTCAGACCGCTGAGCAACATGGCAAAACCCCTGTTTCTATA-----ACAAAA 3681
Db      25      ArgSerSerArgProAlaTrpAlaThr*****AsnProValSerThrLysAsnThrLys 44

```



```
CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS.
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14572; -; NOT_ANNOTATED_CDS.
DR DOMAIN 1 97 FRAME-1.
KW Hypothetical protein.
FT DOMAIN 101 196 FRAME-2.
FT DOMAIN 200 295 FRAME-3.
FT DOMAIN 299 395 FRAME-4.
FT DOMAIN 399 494 FRAME-5.
FT DOMAIN 498 593 FRAME-6.
```

```
SQ SEQUENCE 593 AA; 64603 MW; 136EF344AACD12A2 CRC64;
Alignment Scores:
Pred. No.: 1,8e-45 Length: 593
Score: 526.00 Matches: 196
Percent Similarity: 34.49% Conservative: 33
Best Local Similarity: 29.52% Mismatches: 129
Query Match: 5.45% Indels: 308
DB: 1 Gaps: 18

US-09-966-880A-9 (1-5514) x ALU6_HUMAN (1-593)
QY 3568 TGGCTCAGCGCTGTAATCCAGCAGCTCTGGGAGGCTGAGGTGGTGCCTTGCAGGCC 3627
Db 5 TrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlySerProGluVal 24
QY 3628 TGGAGTTCAAGCAGCCTGAGCAACATCGGCAAAACCTGTTCTCTATA-----ACAAA 3681
Db 25 GlySerSerArgProAla***ProThrTrpArgAsnProValSerThrLysAsnThrLys 44
QY 3682 ATTACCGCGGCGATGTGTCATGTGCTGTGCTGCCAGCTACTAGGGG-GCTGAGGCAGA 3740
Db 45 IleSerArgAlaIleTrpArgMetProValIleProAlaThrArgGluAlaGluAlaGly 64
QY 3741 GAATCTTTGGAGCCGAGGAGGTCAAGGCTGAGCAGTGCCTTGCAGCCACTGCACCTCC 3800
Db 65 GluSerLeuGluProGlyArgArgArgLeuArg***AlaGluIleAlaProLeuHisSer 84
QY 3801 AGCCTGGGTGACAGCAGCAGCAGCTTGCCTCAAAAATAAGAAATAAATAAATAAATAA 3860
Db 85 SerLeuGlyAsnLysSer-GluThrProSerGlnLys-----96
QY 3861 TGGAAACAACATACAAAGAGCTGTTGCTCTAGATGAGCTACTTAGTTAGGTCATATTG 3920
Db 96-----96
QY 3921 GTATTTAACTTTTAAAGTCAGGGTCTGTCACTGCACCTACATATTATTAATAATCAATTCT 3980
Db 96-----96
QY 3981 CAATGTATATCCACACAAAGACTGGTCAATGTTTCATAGTACCTTTATTCACAAAC 4040
Db 96-----96
QY 4041 CCCAAAGTAGAGACTATCCAAATATCCATCAACAAGTCAACAATAAACAATAATCGCTA 4100
Db 96-----96
QY 4101 TATCCATGCAATGGAATACCACCTGTCAGTACAAAGGAAGAAGCTACTTGGGATGAATC 4160
Db 96-----96
QY 4161 CCAAGTCATCAGCGCTAAATCAAGAGTCACACATGAAGAGGAGAGATTAATGTATGCCATA 4220
Db 96-----96
QY 4221 CGAAATCTCAGAAATGAAAGTAAGTTATAGTTACAGAAAGCAAAATCAGGCGAGCATAG 4280
Db 97-----Lys*****AlaGlyArgG1 104
QY 4281 AGGCTCAGCAGCTGTAATCCAGCAGCTTTGAGAGGCGCAGCTGG-GAAGATTGCTAGAACTC 4339
Db 104 yGlySerArgLeu***SerGlnHisPheGlyArgProArgAlaAspHisLeuArgSe 124
QY 4340 AGGAGTTCAAGCAGCAGCTGGGCAACACAGTGAAGAACTCCATCTCCACAAATAATGGGAAA 4399
Db 124 rGlyValArgAspGlnProAspGlnHisGlyGluThrPro-SerLeuLeuLysIleGlnL 144
QY 4400 AAAAGAAAGCAAAATCAGTGGTGTCTGCTGGGAGGGAAGAGACTGCAAGAGGGAAGAA 4459
Db 144 ys-----LeuAlaGlyArgGlyGlyAlaCysLeu***SerGlnL 157
QY 4460 GCTCTGTGGGTGGGTGAGGGTGTGATTGATTGATTGATTGATTGATTGATTGATTGATTG 4519
```


CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

DR EMBL; U14574; -; NOT_ANNOTATED_CDS.
 KW Hypothetical protein.
 FT DOMAIN 1 96 FRAME-1.
 FT DOMAIN 100 195 FRAME-2.
 FT DOMAIN 199 294 FRAME-3.
 FT DOMAIN 298 393 FRAME-4.
 FT DOMAIN 397 492 FRAME-5.
 FT DOMAIN 496 591 FRAME-6.
 SQ SEQUENCE 591 AA: 64395 MW: AC8154AD8A6BB280 CRC64;

Alignment Scores:

Pred. No.: 4, 14e-45 Length: 591
 Score: 522.50 Matches: 215
 Percent Similarity: 34.36% Conservative: 54
 Best Local Similarity: 27.46% Mismatches: 173
 Query Match: 5.41% Indels: 342
 DB: 1 Gaps: 24

US-09-966-880A-9 (1-5514) x ALU8_HUMAN (1-591)

QY 3552 AGACAGCCAGGTGGTGGCTCACGCTGTAATCCAGCAGCTCTGGAGGCTGAGGTGG 3611
 Db 3 ArgArgProGluAlaValAlaHisAlaCysAsnProSerThrLeuGlyGly-Arg----- 20
 QY 3612 TGGATCGTTGAGCCCTGGAGTTCAGACCCAGCCGTGAGCAACATGGCAAAACCCCTGTTTC 3671
 Db 21 ----SerProGluValArgSerArgProAlaThrProThrTrp***AsnProValSe 39
 QY 3672 TATA-----ACAAAAATTTAGCCGGCATGGTGGCATGGCTGGTCCAGCTACTAG 3725
 Db 39 rThrLysAsnThrLysIleSerArgAlaThrProAlaThrProAlaThrProAlaThrAr 59
 QY 3726 GGG-GCTGAGGAGGAGATCTTTGGAGCCAGGAGGTCAAGGTCAGCTGACAGCTGCT 3784
 Db 59 gGluAlaGluAlaGlyGluSerLeuGluProGlyArgArgArgGln***AlaGluI 79
 QY 3785 TCGGCCACTGCACTCAGCCTGGGTGACAGGACGACCTTCGCTCAAAAAATAGAG 3844
 Db 79 eAlaProLeuHisSerSerLeuGlyAspArgAlaArgLeuArgLeuLysLys----- 96
 QY 3845 AAAAAATTAATAATGAAACAACTACAAAGAGCTGTGCTAGATGAGCTACTTAG 3904
 Db 96 ----- 96
 QY 3905 TTAGCTGATATTTTGGTATTAACTTTTAAAGTCAGGCTCTGTCACCTGCACATTA 3964
 Db 96 ----- 96
 QY 3965 TTAATAATCAATTCATGTATATCCACAAAGACTGGTACGTGAATGTTTCATAGTA 4024
 Db 96 ----- 96
 QY 4025 CCTTTATTCAAAACCCCAAGTAGAGACTATCCAAATATCCATCAACAGTGAACAAA 4084
 Db 96 ----- 96
 QY 4085 TAACAAAAATGTCTATATCCATGCAATGGAATACCACCTGCAGTACAAAGGAAGC 4144
 Db 96 ----- 96
 QY 4145 TACTTGGGATGAATCCCAAGTCATGACGCTAAATGAAGAGTCAGACATGAAGGAGGA 4204
 Db 96 ----- 96
 QY 4205 GATAATGTATGCCATACGAAATTCATAGAAAATGAAAGTAACATTATAGTTACAGAAAGCAA 4264
 Db 97 -----*****AlaGI 101
 QY 4265 ATCAGGCGCATAGAGGCTCACACCTGTAATCCAGCAGCTTTGAGAGGCCACGTGGGA 4324

Db 101 yGlyGlyArgAlaArgTrpLeuThrProValIleProAlaLeuTrpGlu-----GI 118
 QY 4325 AGATTGCTAGAACTCAGGAGTTCAGACCCAGCTGGGCAACACAGTGAACCTCCATTCTC 4384
 Db 118 uAspHisLeuArgSerGlyValArgAspGlnProGlyGlnHisGlyGluThrProSerLe 138
 QY 4385 CACAAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTGTCTCTGTGG----- 4430
 Db 138 uLeuLysIleGlnLysLeuAlaGlyArgGlyAlaArgLeu***SerGlnLeuLeuGI 158
 QY 4431 -----GGAGGGGAAGGACTGCAAGAGGAGGAGAA 4459
 Db 158 yArgLeuArgGlnGluAsnArgLeuAsnProGlyGlyGlyGly-CysSerGluProArg 178
 QY 4460 GC----- 4461
 Db 178 erArgHisCysThrProAlaThrProAlaThrGluArgAspSerValSerLysLys***** 198
 QY 4462 -----TCTGTGGGTGAGGTGGTGTGATTCAG----- 4488
 Db 198 **ArgAlaGluAlaGlyArgGlyGlySerArgLeu***SerGlnHisPheGlyArgLysI 218
 QY 4489 --GTTCTGTATCTGACGTGTGTGAGTGTG-----GGTGTCTTACATCCAAATATTCG 4543
 Db 218 leThr**GlyGlnGluPheGluThrSerLeuAlaAsnMetValLysProArgLeuTr* 238
 QY 4544 TAGAATTATGATCTTAATGTTGAGTGTGAGTGTGATGATGATTAATATCTCAATCTAAG 4603
 Db 238 **Lys-----TyrLysAsn***ProGlyValVala 248
 QY 4604 AAAAAATATGTGAAGAAAGTCAATCTCTTCCAGCAAACTTATCAATCTCCT 4663
 Db 248 laArg---AlaCysAsnProSerTrpSerGlyGly**GlyArgArgIleAla***Thra 267
 QY 4664 GAGCCTTCTACTTCCCA-----AATTCTCTGACCTCTGCCCGTACCATTAG 4711
 Db 267 rgLuuAlaGluValAlaValSerArgAspArgAlaThrAlaLeuGlnProGlyArgGlns 287
 QY 4712 GTGACAGCACTAGCTCCCAAAATGGATAAATGCATTTCTGGAAAAAGACTAGGACAAAA 4771
 Db 287 erGluThrProSerGlnLysLys*****PhePheLeuArgArg----- 302
 QY 4772 TCCAGCATCAGCTGTGCTTTCATATCAACACCGTGTACAGCTGTGTGTCTGTCTGCA 4831
 Db 303 -----SerLeuAlaLeuSerProArgLeuGluCysSerGlyAlaIleSer--Alah 319
 QY 4832 GGTGCAAT-----GGGACTCT----- 4848
 Db 319 isCysAsnLeuArgLeuProGlySerSerAspSerProAlaSerAlaSerArgValAlaG 339
 QY 4849 -----TGATTCTTTAAGGAACTGG 4870
 Db 339 lyIleThrGlyAlaArgHisAlaArgLeuIlePheValPheLeu---ValGluThrG 358
 QY 4871 GTTACAGAGTATTTCCAAATGCTATTCAA---ATTAGTCTTATGATATCAACAGACA 4927
 Db 358 lyPheHisHisValGlyGlnAlaGlyLeuGluLeuThrSerGlyAspLeuProL 378
 QY 4928 CTGTGCTAGGA----- 4938
 Db 378 ysValLeuGlyLeuGlnAla***AlaThrAlaProGlyLeuArgPro*****PheP 398
 QY 4939 -----GCCAGAAAAACAAGAG 4954
 Db 398 he***AspGlyValSerLeuCysArgProGlyTrpSerAlaValAlaArgSerArg---- 416
 QY 4955 GAGGAGAAATCAGTATTATGTGGGAACAACATAGCAAGATATTATAGATCATTTTGACTA 5014
 Db 417 -----LeuThra 419
 QY 5015 GTTAAAAAGCAGCAGAGTACAAAATCACAATGCAATGCAATATCCAAATCATGTAA 5074
 Db 5015 ----- 5074

potentially misleading protein sequences.";
Genomics 12:838-841(1992).
[3]
ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
J. Mol. Evol. 27:194-202(1988).
[4]
ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
J. Mol. Evol. 32:105-121(1991).
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U14573; -; NOT_ANNOTATED_CDS.

DR Hypothetical protein.
KW DOMAIN 1 97 FRAME-1.
FT DOMAIN 101 196 FRAME-2.
FT DOMAIN 200 295 FRAME-3.
FT DOMAIN 299 395 FRAME-4.
FT DOMAIN 399 494 FRAME-5.
FT DOMAIN 498 593 FRAME-6.
SQ SEQUENCE 593 AA; 64417 MW; 54A4F50F33A6089F CRC64;

Alignment Scores:
Pred. No.: 1.05e-41 Length: 593
Score: 489.50 Matches: 200
Percent Similarity: 36.28% Conservative: 50
Best Local Similarity: 29.03% Mismatches: 187
Query Match: 5.08% Indels: 234
DB: 1 Gaps: 20

US-09-966-880A-9 (1-5514) x ALU7_HUMAN (1-593)
QY 5512 CTGAGTCTAGGAGTTTGAGACCGCTGGCCACATGGTGAACCCCATCTCTACTAATAA 5453

122 LeuArgSerGlyValArgAspGlnProGlyGlnHisGlyGluThrProSerLeuLeuLys 141
5452 ATACAAACATTAGCCGAGTGTGGTGGCGGCTGTAAATCCCATCTACCCGGGAGGCTG 5393
142 IleGlnLysLeuAlaGlyArgGlyGlyArgLeu***SerGlnLeuLeuGlyArgLeu 161
5392 AGCGCAGAGCTTTGCTTGAACCCAGAGGTGGAGTTGCAGTGAGCTATGTCATGCAT 5333
162 ArgGlnGluAsnArgLeuAsnProGlyGlyGlyCysSerGluProArgSerArgHis 181
5332 TCCATCCAGCATGGGCAACAGCAAACTCCATCTCAAAAAAATAAAAAAATAAAAA 5273
182 CysThrProAlaTrpAlaThrArgAlaLysLeuArgLeuLysLys*****Pro--- 200
5272 AAAAGAAAGAGTAATCTTGCATCTCTCAAACTTGAATCTATGGAATAATAGACAGC 5213
200 -----
5212 TCTAAATTTTAAATATTAAGAGATTTTAAATCTCAAAATATCATAGTGTCTAATAATA 5153
200 -----
5152 TAATTAGACTTACTGTCTAATGACAATGACTGTTAAGATTCTTGTGTATTCTCTCTAGT 5093
201 -----
5092 CTTTCTACA-----GGCACATATTACATGATTGGATTACTGATTGC 5048
209 ProSerThrLeuGlyGlyArgGlyGlyTrpIleThr***GlyGlnGluPheGluThrSer 228
5047 ATGTGTGATTTTGTACTCTGCTCTTTTAACTAGTCAAAATGATCTAAATATCTTGCT 4988
229 LeuAlaAsnMetVal-----LysProArgLeuTyr***--- 239
4987 ATGTTGTTCCACATATGACTGATTTCTCTCTCTCTTTGTTGCTGCTCTAGACACAG 4928
239 -----
4927 TGTCTTGCATATCATAGCACTAATTTGAATAGCATTTGTGGAAATACTCTGCTAACCCA 4868
240 -----LysTyrLysAsn***Pro 245
4867 AGTTTCTTAAAGAAATCAAGAGTCCCATTCGAGCTGCAGACAGCAACAAAGCTGTAC 4808
246 GlyValValAlaGlyAlaCysAsnProSerTyrSerGlyGly***GlyArgArgIleAla 265
4807 AGCGTGTTCATATGAAGACACAAAGTGCCTGGATTTTGTCCCTAGTCTTTTCCAGAA 4748
266 ***ThrArgGluAlaGluValAlaVal-----
4747 ATGCATTTATCCAAATTTGTGGAGCTAGTGTCTACCTAATGTCACGGGCGCAGAAGTGCA 4688
275 -----SerArgAspArg 278
4687 GAGAAATTCGGAAGTAAAGGGCTCAGGAATTTGAATACGTTTGTGGCAAGAGAAATTGA 4628
279 AlaThrAlaLeuGlnProGlyGlnGluArgAsnSerVal-----
4627 AACTTTTCTTACATATTATTTTCTTACATGAGGTATAATTTTACATACAGTAAGTCC 4568
293 SerLysLys*****PhePheLeuArgArgSerPheAlaLeuValAlaGlnAlaGly 312
4567 ACCCATTTTAAGATGCATAATTTCTACGAATATTTTGGATGTTAAACACCCCAAACTCCTAC 4508
313 ValGlnTrpArg-----
4507 CACAGTCAGGATACAGAACCTGAATCACCACCTCACC-----CACACAGAGCTTCTTCCCT 4451
317 -----AspLeuGlySerLeuGlnProProGlyPheLysArgPheSerCys 333
4450 CTTTGCAGCTCTTCC-----CCTCCCCCAGGACACCACTGATTTG 4409

```
Db 334 LeuSerLeuProSerSerTrpAspTyrArgArgProProArgProAlaAsn----- 351
Oy 4408 CTTTCTTTTTCCTCCATTTTGTGGAGATGGAGTTCTCACTGTCTGTCGCCAGGCTGTCT 4349
Db 352 -----PheCysIlePheSerArgAspGlyValSerProCysTrpProGlyTrpSer 368
Oy 4348 TGAACCTCTGAGTCTTACGAATCTCCACGCT-GGCCTCTCAAAAGTCTGGGATTACAGG 4290
Db 369 ArgThrProAspLeuArg***SerThrArgLeuGlyLeuProLysCysTrpAspTyrArg 388
Oy 4289 TGTGAGCCT-----CTATGCGCTG 4272
Db 389 ArgGluProProArgProAla*****PhePhe***AspGlyValSerLeuLeu 408
Oy 4271 CCCTGATTT-----GCTTTCGTACTATAGTTACTTTCATTT 4233
Db 409 ProArgLeuGluCysSerGlyAlaIleSerAlaHisCysAsnLeuArgLeuProGlySer 428
Oy 4232 TCTAGAATTTTCGTATGCATACATTATCTCCTCTTCATCTCTGCTCTTCATTTAGCG 4173
Db 429 Ser----- 429
Oy 4112 ATTCATGGATATAGCACATTTTGTATTGTTTCACTTGTGTGATGATATTTGGATAGT 4053
Db 442 rGlyAlaArgHisAlaArgLeuIlePheValPheLeuVal----- 456
Oy 4052 CTCTACTTTGGGTTTGTGAATAAGTACTATGAACATTCACGTACCGTCTTTGTGT 3993
Db 457 -----GluThrGlyPheHisValGlyGlnAlaGlyLeu----- 468
Oy 3992 GGATATACATGAGATGATATTTTAAATAGTAGTCAGGTGACAGACCCCTGACTTA 3933
Db 469 -----GluLeuLeu-----ThrSerGlyAspProPro----- 477
Oy 3932 AAAGTTAAATACCAAAATATACGCTAACTAAGTAGCTCATCTAGGACACAGCTCTTTG 3873
Db 478 -----AlaSerAlaSerGlnSerAlaGlyIleTh 487
Oy 3872 TAGTGTGTTTCATTTTATTTTAAATTTTCTTCTTATTTTGTAGGCAAGCTCGTCT 3813
Db 487 rGly-ValSer---HisArgAlaArg*****PhePheGluThrGluPheArgSerC 506
Oy 3812 GTCACCCAGGCTGAGTGTCAGTGGCGAAGCACTGCTCAGTGCAGCCTTGACCTCCTGGG 3753
Db 506 ysCysProGlyTrpSerAlaValAlaArgSerArgLeuThrAlaThrSerAlaSerArgV 526
Oy 3752 CTCCAAGATCTCTCCCTCAGCCCTAGTAG-CTGGGACCACAGGCACATGCCACCA 3694
Db 526 alGlnAlaIleLeuLeuProGlnProProGlu***LeuGlyLeuGlnAlaProAlaThr 546
Oy 3693 TGCCCGGCTAATTTTG-----TATAGAAACAGGGTTTGGCATGTGTCTGAGGCTGG 3640
Db 546 hrProGly**PheLeuTyrPhe*****ArgArgGlyPheThrMetLeuAlaArgLeuV 566
Oy 3639 TCTTGAACCTCAGGGCTCAAGCGATCCACCACCTCAGCCTCCAGAGTGTGGGATTAC 3580
Db 566 alSerAsnSer***ProGlnValIleHisProProArgProProLysValLeuGlyLeuG 586
Oy 3579 AGCGGTGAGCCACACCTGGC 3557
Db 586 InAla***AlaThrAlaProGly 593

RESULT 7
ALU8_HUMAN
ID ALU8_HUMAN STANDARD; PRT; 591 AA.
AC P39195;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
```

Alu subfamily SX sequence contamination warning entry.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=95021758; PubMed=7935834;
Claverie J.-M., Makalowski W.;
"Alu alert."
Nature 371:752-752(1994).
[2]
CONCEPT.
MEDLINE=92241891; PubMed=1572661;
Claverie J.-M.;
"Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
Genomics 12:838-841(1992).
[3]
ALU FAMILIES CLASSIFICATION.
MEDLINE=88333009; PubMed=3138422;
Quentin Y.;
"The Alu family developed through successive waves of fixation
closely connected with primate lineage history.";
J. Mol. Evol. 27:194-202(1988).
[4]
ALU FAMILIES CLASSIFICATION.
MEDLINE=91178815; PubMed=1706781;
Jurka J., Milosavljevic A.;
"Reconstruction and analysis of human Alu genes.";
J. Mol. Evol. 32:105-121(1991).
-1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
REPEATS.
-1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CODON. 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
-1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
ACID SEQUENCES.
-1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CNAS
LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
BEING REPORTED.
-1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CODING NUCLEOTIDE SEQUENCE.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.

EMBL; U14574; ; NOT_ANNOTATED_CDS.
Hypothetical protein.
DOMAIN 1 96 FRAME-1.
DOMAIN 100 195 FRAME-2.
DOMAIN 199 294 FRAME-3.

FT DOMAIN 298 393 FRAME-4.
FT DOMAIN 397 492 FRAME-5.
FT DOMAIN 496 591 FRAME-6.
SQ SEQUENCE 591 AA: 64395 MW: AC8154AD8A6BB280 CRC64;

Alignment Scores:
Pred. No.: 1.66e-38 Length: 591
Score: 458.50 Matches: 191
Percent Similarity: 37.11% Conservative: 68
Best Local Similarity: 27.36% Mismatches: 160
Query Match: 4.76% Indels: 280
DB: 1 Gaps: 22

US-09-966-880A-9 (1-5514) x ALU8_HUMAN (1-591)

QY 5512 CTGAGTCTAGGAGTTTGACACAGCCTGCCACATCGGTGAACCCCATCTCTACTAAAA 5453
DB 121 LeuArgSerGlyValArgAspGlnProGlyGlnHisGlyGluThrProSerLeuLeuLys 140
QY 5452 ATACAAACATTAGCCGAGTGTGGTGGCGGCTGTAATCCCATCTACCCGGGAGGCTG 5393
DB 141 IleGlnLysLeuAlaGlyArgGlyGlyAlaArgLeu***SerGlnLeuGlyArgLeu 160
QY 5392 AGCGACAGCTTTGTGAACCCAGGAGGTGGAGTTGGAGTGAGTATGRTCATGCCAT 5333
DB 161 ArgGlnGluAsnArgLeuAsnProGlyGlyGlyCysSerGluProArgSerArgHis 180
QY 5332 TCCACTCCAGCATGGCCACAGACCAAACTCCATCTCAAAAAAATAAAAAA 5273
DB 181 CysThrProAlaTrpAlaThr-GluArgAspSerValSerLys*****ArgAl 200
QY 5272 AAAAA-----GAAAGAGTAATACTTGATCATCTGTCAAACTTGAATACATG 5228
DB 200 aGluAlaGlyArgGlySerArgLeu***SerGlnHisPheGlyArg----- 216
QY 5227 GAAATAGAGAGCTCTAAATTTTAAATATTAAGATTTTAAATCTCAATATCAT 5168
DB 217 -LysIleThr***GlyGlnGluPheGluThrSer-----LeuAlaAsnMetVa 232
QY 5167 AGTGCTCTAATAATAATAGACTTAGTGTCTAATGACAATGACTGTTAAGATCTTGT 5108
DB 232 I----- 232
QY 5107 GTTATTCTCTAGTCTTTCTACAGGCACATATTTACATGATTGG----- 5062
DB 233 -----LysProArgLeuTyr***LysTyrLysAsn***ProGl 245
QY 5061 -ATTATAGTATGCATGTGTGATTTTGTACTCTGCTGCTTTT-----TTACTAG 5012
DB 245 yValValAlaArgAlaCysAsnProSerTyrSerGlyGly***GlyArgArgIleAla** 265
QY 5011 TCAAAATGATCTAAATATCTGCTATGTTGTTGCCACATATGACTGATTCTCCCTC 4952
DB 265 *ThrArgGluAlaGluValAla-----ValSerArgAspArgAlaThrAl 280
QY 4951 TTTGTTTCTGGTCTCTAGCACAGTGTCTTGCATATCATAAAGCACTAATTGAATAGCAT 4892
DB 280 aLeuGlnProGly----- 284
QY 4891 TTGTGGAATACTCTGGTAACCCCAAGTTTCTTTAAAGAAATCAAGAGTCCCATTCGAGC 4832
DB 285 -----ArgGlnSer-GluThrProSerGlnL 293
QY 4831 TGCAGACAGCAACACAGCTGTACAGCGTGGTGTATGAATGAAGCAAGTGATGCCTGGA 4772
DB 293 ysllys*****PhePheLeuArgArgSer----- 303
QY 4771 TTTTGTCCCTAGTCTTTTCCAGAAATGCATTTATCCAAATTTGTGGAGCTAGTCTGTCAC 4712
DB 304 --LeuAlaLeu-Ser---ProArgLeuGluCysSerGlyAlaIleSerAlaHisCysAsn 321
QY 4711 CTAAATGGTACGGGCAAGATGCAGAGAATTTGCGAAGTAAGGGCTCAGGAATTGAAAT 4652
DB 321 I--- 321

DB 322 LeuArgLeuProGlySerSer----- 328
QY 4651 AACGTTTGTGCGCAAGAGAATTGAAACTTTTCTTACACATTAATTTTCTTACATTGAGG 4592
DB 328 ----- 328
QY 4591 TATAATTTTACATACAGTAAACTCCACCCATTTTAAGATGCATAAATTTCTACGAATATTTTGTG 4532
DB 328 ----- 328
QY 4531 GATGTAAACACCCCAAACTGCTTACCACAGTCAGGATACAGAACCTGAATCACCACCTCA 4472
DB 329 -----AspSerProAlaSer 333
QY 4471 CCCCACAGAGCTTCTCCCTCTTTTGCAGTCTTCCCTCCCCACAGGACAACCACTGAT 4412
DB 334 -AlaSerArgValAlaGlyIle-----ThrGlyAlaArgHisH1 346
QY 4411 TTGCTTTCTTTTTCCTTTTCCCATTTTGTGGAGAAATGGAGTTTCTGTTTCCCGAGGCTGG 4352
DB 346 sAlaArgLeuIlePheValPheLeuValGluThrGlyPheHisValGlyGlnAlaG1 366
QY 4351 TCTTGAACCTCTGAGTCTTAGCAATCTCCACAGTGGCCTCTCAAAAGTGTCTGGATTACA 4292
DB 366 yLeuGluLeuLeuThrSerGlyAspLeuProPro-----LysValLeuGlyLeuG1 383
QY 4291 GGTGTGAGCCTCTATGCTGCTGCTGCTTCTTCTGTAACTATAAGTTACTTTTCATTTT 4232
DB 383 nAla***AlaThrAlaPro----- 389
QY 4231 CTAGAATTTGCTATGGCATACATATCTCTCTCTCATGTCTGACTCTTTTCATTTACCGT 4172
DB 389 ----- 389
QY 4171 CATGACTTTGGGATTTCATCCCAAGTAGCTTCTTC-----CTTTGTA 4130
DB 390 -----GlyLeu-ArgPro*****PhePhe***AspGlyValSerLeuCysA 406
QY 4129 CTGACAGGTGTTATTCATTCATGATGATATAGCACATTTTGTATTATTTGTTCACTTGTG 4070
DB 406 rgProGlyTrpSerAlaValAla----- 413
QY 4069 ATGGATATTTGGATAGTCTCTACTTTGGGGTTTGTGAATAAGGTACTATGAACATTC 4010
DB 414 -----ArgSerArgLeuThrAlat 420
QY 4009 CGTACCAGTCTTTGTGTGGATATATATGAGAATTTGATATTTTAATATGTTAGTGCAGGT 3950
DB 420 hrSerAlaSer-----ArgValGlnAlaIleLeuLeuProGlnProp 434
QY 3949 GACAGACCTCTGACTTTTAAAGTTAAA----- 3924
DB 434 roGlu***LeuGlyLeuGlnAlaArgAlaThrThrProGly***PheLeuTyrPhe*** 454
QY 3923 --TACCAAAATATCAGCCTAACTAAGTAGTCTATCTAGGACACAGCTCTTTGTAGTTGT 3866
DB 454 **ArgArgGlyPheThrMetLeuAlaArgLeuValSerAsnSer***ProGlnValIleP 474
QY 3865 TT---CCATTTTATTTTAAATTTT-----TCTT 3842
DB 474 heLeuProLysCysTrpAspTyrArgArgGluProProArgProAlaSerAlaArg*** 494
QY 3841 CTTATTTTGTGGCAAGGTCTGCTGCTCTGTACCCAGGTGGAGTGCAGTGGCGGCAAGC 3782
DB 494 ****PhePheGluThrGluSerArgSerValAlaGlnAlaGlyValGlnTrpArgAspL 514
QY 3781 ACTGCTCAGTCGACGCTTGACCTCTGGCTCCCAAGATTTCTCTGCTCAGCCCC-CCTA 3723
DB 514 euGlySerLeuGlnProProProGlyPheLysArgPheSerCysLeuSerLeuProS 534
QY 3722 GTAGCTGGGACCAACAGCAGCATGCCACCATGCCCGCTAAATTTTGTATT-----AGAA 3669
DB 534 erSerTrpAspTyrArgArgAlaProProArgProAlaAsnPheCysIlePheSerArgA 554


```
Db 252 ----- 252
QY 4912 AAGCACTAATTGTAATAGCATTGTGGAAATCTCTGGTAACCAAGTTTCCCTAAAGAA 4853
Db 253 -----AsnProSerTyrSerGlyGly 259
QY 4852 ATCAAGAGTCCCATTCAGCTGCAGACAGCAACACAAGCTGTACACGCTGGTGTATATG 4793
Db 260 **GlyArgArgIleAla**ThrArgGluAlaGluValSer---ArgAspArg 278
QY 4792 AAAGCACAAGTGCCTGGATTTGTCCTAGTCTTTTCCAGAAATGCATTTATCCAAT 4733
Db 279 AlaIleAlaLeuGlnPro----- 284
QY 4732 TTGTGGAGCTAGTGCTGTACCTAATGTTACGGGGCAGAGTGCAGAGAAATTTGCGAAGT 4673
Db 285 -----GlyGlnGlnArgAsnSerValSer 293
QY 4672 AAAGGCTCAGGAATTTGAATAACGTTTGTGCGCAAGAGAATTTGAAACTTTTCTTACACA 4613
Db 294 Lys-----Lys***** 297
QY 4612 TATATTTTCTTACATTGAGGTATATTTACATACAGTAACTCCACCCATTTAAGATGC 4553
Db 298 ***PhePheLeuArgArgSerPheAlaLeuValAlaGlnAlaGlyValGlnTrpArg--- 316
QY 4552 ATAATTCTACGAATATTTTGGATGTAAACACACCCCAACTGCTTACCACAGTCAGGATACA 4493
Db 317 -----AspLeuGlySer----- 320
QY 4492 GAACCTGAATACACACCCACCCACAGAGCTTCTCCCTGTTGCGAGTCCTCC--- 4436
Db 321 ---ProGlnProProGlyPheLysArgPheSerCysLeuSerLeuProSerSer 339
QY 4435 -----CCTCCACAGGACACCACTGATTTGCTTCTTTTCTCC 4394
Db 340 TrpAspTyrArgHisAlaProProArgProAlaAsn-----PheCys 353
QY 4393 ATTTTGTGGAGAATGAGTTTCACCTGTGTGCGCCAGGCTGTGAACCTCTCGAGTTC 4334
Db 354 IlePheSerArgAspGlyValSerProCysTrpSerGlyTrpSerArgTrpProAspLeu 373
QY 4333 TAGCAATCTCCACGCT-GGCCCTCAAAGTGTGGATTACAGGTGTGAGCCT----- 4281
Db 374 Arg***SerAlaArgLeuGlyLeuProLysCysTrpAspTyrArgArgGluProProArg 393
QY 4280 -----CTATGCTCGCCTGATTTGCTTTC 4257
Db 394 ProAla*****PhePhe***AspGlyValSerLeuLeuLeuProArgLeuGlu--- 412
QY 4256 TGTAACTATAAGTACTTTCTCATTTTCTAGAATTTGCTATGGCATACATTATCTCCCTTC 4197
Db 413 CysAsnGlyAlaIleSerAlaHisArgAsnLeu-ArgLeu-----ProG1 427
QY 4196 CATGCTGACTCTTTCATTTAGCGTCATGACTTTGGGATTCATCCCAAGTACTCTTTC 4137
Db 427 ySerSerAspSer-----ProAlaSerAlaSerAr 437
QY 4136 CTTTGTACTGCAGGGTGTATTCATTGCATGATATAGCACATTTTGTATTATTGTCA 4077
Db 437 gValAlaGlyIleThrGlyMetArgHisAlaArgLeuIle-PheValPheLeuValG 457
QY 4076 CTTGTTGATGGATATTGGATAGTCTCTACTTTGGGGTTTGTGAATAAAGGTACTATGA 4017
Db 457 luthr-----GlyPheLeuHisValGlnAlaG 467
QY 4016 ACATTACGTACCAGCTTTTGTGTGTGATATACATTGAGAATTATTTTAATAAGTAG 3957
Db 467 lLeuGluLeuProThr----- 472
QY 3956 TGCAGGTGCAGACACCTGACTTTTAAAGTTTAAATACAAATATACAGCTTAAGTAG 3897
Db 473 ----SerGlyAspProPro-----A 478
```

```
QY 3896 CTCACTAGGACAACAGCTCTTTGTAGTGTCTTCCATTTATTTTAAATTTTCTTCTAT 3837
Db 478 laSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArg*****P 498
QY 3836 TTTTGTGAGCAAGGTCTGTCTGTCAACCCAGGCTGGAGTGGCGCAGCAAGCACTGC 3777
Db 498 hePheGluThrGluPheArgSerCysCysProGlyTrpSerAlaMetAlaArgSerArgL 518
QY 3776 TCAGTGCAGCCTTGACCTCTCTGGCTCCAAAGATTCTCTGCTCAGCCCTAGTAG-C 3718
Db 518 euthrAlaThrSerAlaSerArgValGlnAlaIleLeuLeuProGlnProGlu***L 538
QY 3717 TGGGACCAACAGGCATGCCACCATCCCGGCTAATTTTGTG-----TTATAGAAACAGG 3664
Db 538 euGlyLeuGlnAlaCysAlaThrThrProGly***PheLeuTyrPhe***ArgArgG 558
QY 3663 GTTTTGCATGTTGCTCAGGCTGTGTTGAACCTCCAGGCTCAAGCGCATCCACCACCTC 3604
Db 558 lypheSerMetLeuValArgLeuValSerAsnSerArgProGlnValIleArgProProA 578
QY 3603 AGCCTCCAGAGTGTGGGATTACAGGCTGAGCCACACACCTGGC 3557
Db 578 rgProProLysValLeuGlyLeuGlnAla***AlaThrAlaProGly 593

RESULT 9
ALU5_HUMAN STANDARD; PRT; 585 AA.
ID ALU5_HUMAN AC P39192;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SC sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
REPEATS.
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
ACID SEQUENCES.
```

CC -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
 CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
 CC LOCATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
 CC BEING REPORTED.

CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
 CC CODING NUCLEOTIDE SEQUENCE.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: U14571; ; NOT_ANNOTATED_CDS.
 CC KW Hypothetical protein.
 CC FT DOMAIN 1 95 FRAME-1.
 CC FT DOMAIN 99 193 FRAME-2.
 CC FT DOMAIN 197 291 FRAME-3.
 CC FT DOMAIN 295 389 FRAME-4.
 CC FT DOMAIN 393 487 FRAME-5.
 CC FT DOMAIN 491 585 FRAME-6.
 CC SQ SEQUENCE 585 AA; 63957 MW; 46EE8C4F93650A7 CRC64;

Alignment Scores:

Pred. No.: 1.01e-35 Length: 585
 Score: 431.50 Matches: 187
 Percent Similarity: 34.05% Conservative: 49
 Best Local Similarity: 26.98% Mismatches: 187
 Query Match: 4.47% Indels: 271
 DB: 1 Gaps: 19

US-09-966-880A-9 (1-5514) x ALU5_HUMAN (1-585)

QY 3568 TGGCTCAGCGCTGTAAATCCAGCACTCTGGGAGGCTGAGGTGGGTGGATCGCTTGAGCCC 3627
 Db 5 TrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlySerArg--GlyG 24
 QY 3628 TGGAGTTCACAGCAGCCTGAGCAACATGGCAAAACCCCTGTTCTCTATAC-----AAAA 3681
 Db 24 InGluIleGluThrIleLeuAlaAsnMetValLysProArgLeuTrp***LysTyrLysA 44
 QY 3682 ATTAGCCGGGATGTGGTCATGTGCTGTGGTCCAGCTACTA-GGGGGCTGAGCAGGA 3740
 Db 44 sn**LeuGlyValValAlaArgAlaCysSerProSerTyrSerGlyGly***GlyArgA 64
 QY 3741 GAATCTTTGGAGCCAGGAGGTCAAGGCTGCACACTGAGCAGTGCTTGGCCCACTGCACGCC 3800
 Db 64 rgIleAla***ThrArgGluAlaGluValAlaValSerArgAspArgAlaThrAlaLeuG 84
 QY 3801 AGCCTGGGTGACAGCAGCAGCCTTGCCTCAAAAATAAAGA----- 3842
 Db 84 InPro-GlyAspArgAlaArgLeuArgLeuLysLysLys*****AlaGlyArgGlyGly 103
 QY 3843 -----AGAAAAATTAATAATAATGGAACAA 3869
 Db 104 SerArgLeu***SerGlnHisPheGlyArgProArgArgAlaAspHisGluValLysArg 123
 QY 3870 CTACAAAGAGCTGTTGTCTCTAGATGAGCTACTTAGTGGCTGATATTTTGGTATTTAAC 3929

Db 123 ----- 123
 QY 3930 TTTTAAAGTCAGGGTCTGTCACTGCACGTACATATTATTAATAATCAATTTCTCAATGTATA 3989
 Db 124 -----SerArgProSerTrpProThr-Trp***AsnProValSerThrLysAsnThrLys 141
 QY 3990 TCCACACAAAGACTGGTGAATGTTTCATAGTACTCTTTATTTCACAAAACCCCAAGTA 4049
 Db 141 sileSerTrpAlaTrpTrp----- 147
 QY 4050 GAGACTATCCAAATATCCATCAACAAGTGAACAATAAACAATAATGTGCTATATCCATGCC 4109
 Db 148 -ArgAlaProValProAlaThrArgGlu-----AlaGluAl 160
 QY 4110 AATGCAATACCACCCCTGCAGTACAAAGGAAGAGTACTTTGGGATGAATCCCAAGTCA 4169
 Db 160 aGlyGluSerLeuGluProGlyArgArgArgLeuGln***AlaGluIleAlaProLeuH 180
 QY 4170 TGACGCTAAATGAAGAAGTCAGACATGAAGGAGAGATAATGTATGCCATACGAAATTC 4229
 Db 180 sSerSer-----Le 183
 QY 4230 AGAAATGAAGTAACCTTATAGTTACAGAAACCAATCAGGCAGCATAGAGGCTCACA 4289
 Db 183 uAlaThrGluArgAsp-SerValSerLysLys*****ProGlyAlaValAlaHisA 203
 QY 4290 CCTGTAATCCCAAGCAGCTTTGAGAGCCAGCTGGGAAGATTGCTAGAACTCAGAGTTCAA 4349
 Db 203 laCysAsnProSerThrLeuGlyGlyArgGlyGlyArgGlyGlyThrArgSerArgAsp-Arg 222
 QY 4350 GACCAGCCTGGGCAACACAGTGAACCTCCATTCTCCACAAAATGGGAAAAAGAAAGC 4409
 Db 223 AspHisProGlyGlnHisGlyGluThrPro-SerLeuLeuLysIleGlnLys----- 239
 QY 4410 AAATCAGTGGTGTCTCTGTGGGGGGGGAAGGAGCTGCAAGAGGGAAGAGCTCTGGTGG 4469
 Db 240 -----LeuAlaGlyArg-----GlyGly 245
 QY 4470 GGTGAGGTGGTGATTACAGTT-----CTGTATCTCTGA 4502
 Db 245 yAlaArgLeu***SerGlnLeuLeuGlyArgLeuArgGlnGluAsnArgLeuAsnProG 265
 QY 4503 CTGTGTGAGCAGCTTTGGGTGTTTACATCCAAAATATTCGTAGATAATTATGCATCTTAA 4562
 Db 265 yGlyGly-----GlyCysSerGluProArg----- 273
 QY 4563 TGGGTGGAGTTTACTGTATGTAAATATACCTCAATGTAAAGAAAAAATAATGTGTGAAGA 4622
 Db 274 -----SerArgHisCysThrPr 279
 QY 4623 AAGTTTCAATTCCTTTCGCCAGCAACGGTTATTCAAATTCCTGAGCCCTTTACTTCGCAA 4682
 Db 279 oAlaTrpArgGlnSerGluThrProSerGlnLysLys*****PhePheLeuArgar 299
 QY 4683 TTCTCTGCACCTTCTGCCCGCTACCATTTAGGTGACAGCAGCTAGCTCCACAAATTTGGATA 4742
 Db 299 gSerLeuAlaLeuSerPro-----GlyTrpSerAlaValAlaArgSerArgLeuTh 316
 QY 4743 TGCATTTCTGGAAGAAGACTAGGACAAATCCAGGCATCCTGTG----- 4788
 Db 316 rAla-----ThrSerAlaSerArgValGlnAlaIleLeuLeuProGlnProProG 333
 QY 4789 ----CTTTTCATATCAACCCAGCTGTACAGCTTGTGTGTCTGCTGCAGCTGCAATGGGA 4844
 Db 333 u***LeuGlyLeuGlnAlaArgAlaThrThr-----Pr 344
 QY 4845 CTCTTGATTTCTTTAAGGAAACTTGGGTTAC-----CAGAGTATTATTCACAAA 4892
 Db 344 oSer***PheLeu-----TyrPhe*****ArgArgGlyPheThrMetLe 359
 QY 4893 TCGTATTCAAATAGTGTATTGATATGCAAGACACTGTGTAGGAGCCAGAAACAAAG 4952
 Db 359 uAlaArgMetValSerIleSer***ProArgAspProProAlaSerAlaSerGlnSerAl 379

Db 44 sn***ProGlyValValAlaGlyAlaCysSerProSerTyrSerGlyGly***GlyArg 64
 QY 3741 GAATCTTTGGAGCCAGAGCTCAAGCTGACAGCTGCTTGGCCCACTGCACCTCC 3800
 Db 64 rgMetAla***ThrArgGluAlaGluLeuAlaValSerArgAspArgAlaThrAlaLeuG 84
 QY 3801 AGCTGGGTGACAGCAGACCTTGGCTCAAAATAAAGAGAAAAATTAATAAATAA 3860
 Db 84 lnProGlyArgGlnSerGluThr--ProSerGlnLys-----Lys*****Al 100
 QY 3861 TGGAAACAACTACAAAGAGCTGTGTC-----TAGATGAGCTA 3899
 Db 100 aGlyArgGlySerArgLeu***SerGlnHisPheGlyArgProArgArgAlaAspHi 120
 QY 3900 CTTAGTTAGGTGATATTTTGGTATTTAAAGTCAAGGTCTGTCACCTGCACATA 3959
 Db 120 sGluValArg-----ArgSerArgProSerTrpLeuThrAr 132
 QY 3960 CATTATTAATAATCAATCTCAATGTATATCCACACAAAGACTGGTACGTGAATGTCA 4019
 Db 132 g***AsnProValSerThrLysAsnThrLysIleSerArgAlaTrpTrp----- 148
 QY 4020 TAGTACCTTTATTCACAAAACCCCAAGTAGAGACTATCCAAATATCCATCAACAAGTGA 4079
 Db 149 -----ArgAlaProValValProAlaThrArgG 158
 QY 4080 ACAAATAACAAATGTCTATATCCATGCAATGAATACCAACCTGCAGTACAAAGAA 4139
 Db 158 u-----AlaGluAlaGlyGluTrpArgGluProGlyArgArgse 171
 QY 4140 GAAGCTACTTGGGATGAATCCCAAGTCAATGAGCGTAAATGAAAGAGTACAGATGAAG 4199
 Db 171 rLeuGln***AlaGluIleAlaProLeuHisSer----- 183
 QY 4200 GAGGAGATAATGTATGCCATACGAAATCTAGAAAATGAAGTAACTATAGTACAGAA 4259
 Db 184 -----LeuGlyAspArgAlaArgLeuArgLeu-LysL 194
 QY 4260 AGCAATCAGGCGAGGATAGAGCTCACACCTGTATCCAGCAGCTTGGAGGCCAGC 4319
 Db 194 ys*****ProGlyAlaValAlaHisAlaCysAsnProSerThrLeuGlyGlyArg 214
 QY 4320 TGGGAAGATTTCTAGAACTCAGAGTTCAGAACACAGCCTGGCCACACAGTGAACCTCCA 4379
 Db 214 LyGlyArgIleThr-ArgSerGlyAspArgAspHisProGly***HisGlyGluThrPro 233
 QY 4380 TTCTCCCAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTTCCTGTGGGGAGGGAA 4439
 Db 234 SerLeuLeuLysIleGlnLysLeuAlaGlyArg-----GlyGlyGly 247
 QY 4440 GGAAGT-----CAAAGAGGAAGAGCTCTGGTGGGTGAGGTGGTGTATTCAGGTCT 4493
 Db 248 ArgLeu***SerGlnLeuLeuGlyArgLeuArgGlnGluAsnGlyValAsnProGlyGly 267
 QY 4494 GTATCCTGACTGTGTAGTGGTGGGTGTTTACATCCAAATATTCCTGAGAATATG 4553
 Db 268 -----GlyAlaCysSerGluProArgSerArgHis----- 277
 QY 4554 CATCTTAATGGGTGGAGTTTACTGTATGTATTAATATACCTCAATGTAAAGAAAAATAAT 4613
 Db 278 -----CysThrProAlaTrpAlaThrGluArgAspSer 288
 QY 4614 GTGTAGAAAAG---TTTCAATTTCTTGCAGCAAGCTTATTCAAATTCCTGAGCCT 4670
 Db 289 ValSerLysLys*****PhePheLeuArgArgSerLeu-----AlaLeuSerPro 306
 QY 4671 TTACTTCG-----AAATTTCTGCACTTCTGCCCGTACCATTAGGTGACAGCAC 4721
 Db 307 ArgLeuGluCysSerGlyAlaIleSerAla-HisCysLysLeuArgLeuProGlySerAr 326
 QY 4722 TAGTCCCAAAATTTGGATAATGCATTTCTGGAAAAGACTAGGGACAAAATCCAGGCATC 4781
 Db 326 gHisSerProAlaSerAlaSerArgValAlaGlyThrThr-Gly-----AlaArgHisH 344

QY 4782 AC-----TTGTGCTTTTCAT-----ATCAACCACG 4805
 Db 344 isAlaArgLeuIlePheValPheLeuValGluThrGlyPheHisArgValSerGlnAspG 364
 QY 4806 CTGTACAGCTTGTGTGCTGCTGCTGCAATGGGAGCTCTTGTGATTTCTTTAAGGAAA 4865
 Db 364 lyLeuAspLeuLeuThrSer***SerAlaArgLeuGly-----LeuProLysC 380
 QY 4866 CTTGGGTTACCAAGAGTATTTCCA-----CAAATGCTATTCAAATTTAGTGC 4910
 Db 380 yStrpAspTyrArgArgGluProProArgProAla*****PhePhe***AspGlyV 400
 QY 4911 TTATGATATGCAAGACACTGTGCTAGGAGCAGAAACAAAGAGGAGGAGAAATCAGTCA 4970
 Db 400 alSerLeuCysArgPro----- 405
 QY 4971 TTATGTGGAAACAACATAGCAAGATATTTAGATCATATTTGACTAGTTAAAAAAGCAGCAG 5030
 Db 406 -GlyTrp-----SerAlaValAlaArgSerArgLeuThrAlaSerSerAlaSerA 422
 QY 5031 AGTACAAATCACACATCAATCAGTATATATCCAA-----ATCATGTAAATATGTG 5081
 Db 422 rg-----ValHisAlaIleLeuLeuProGlnProProGlu***LeuGlyLeuGlnA 439
 QY 5082 CTTGTAGAAAGACTAGAGGAATAACACAGAATCTTAACAGTCATTGTCTATTAGACACT 5141
 Db 439 laProAlaThrThrProGly----- 445
 QY 5142 AAGTCTAATATTATTATTAGACACTATGATATTTAGATTTTAAATAATCTTTAATATT 5201
 Db 446 -----PheLeuTyrPhe*****ArgArgGlyPheThrValL 459
 QY 5202 TAAATTTTAGAGCTCTCTCTATTTTCCATAGTATTTCAAGTTTGACAATGATCAAGTATTA 5261
 Db 459 eu-----AlaArgMetValSerIleSer***ProArgAspProProAlaSerAlaS 476
 QY 5262 CTCTTTCTTTTCTTTTCTTTT-----TTTTTTTTTTTTTTTGTAGA 5297
 Db 476 erGlnSerAlaGlyIleThrGlyValSerHisArgAlaArg*****PhePheGluT 496
 QY 5298 TGGAGTTTTGTCTTGTTCCTGCTGAGTGGAGTGAATGCATGAYCATGCTACTGCACAA 5357
 Db 496 hrGluSerArgSer-ValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGln 515
 QY 5358 CTCCACCTCTCTGGTTCACAGCAAGCTGTGCTCAGCTCCCGGGGTAGATGGATATAC 5417
 Db 516 AlaProProProGlyPheThrProPheSerCysLeuSerLeuProSerSerTrpAspTyr 535
 QY 5418 AGGCGCCCAACACACACTCGCTGCTCAAACTCCTGACCTC 5510
 Db 536 ArgArgProProProProArgProAlaAsnPheCysIlePheSerArgAspGlyValSerPro 555
 QY 5478 TGTGGCCAGCTGTCTCAAACTCCTGACCTC 5510
 Db 556 Cys***ProGlyTrpSerArgSerProAspLeu 566

RESULT 11
 ALU2_HUMAN
 ID ALU2_HUMAN STANDARD; PRT; 587 AA.
 AC P39189;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alu subfamily SB sequence contamination warning entry.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95021758; PubMed=7935834;
 RA Claverie J.-M., Makalowski W.;

RT "Alu alert."; ;
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RL potentially misleading protein sequences."; ;
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RL closely connected with primate lineage history."; ;
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes."; ;
RL J. Mol. Evol. 32:105-121(1991).
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14568; ; NOT_ANNOTATED_CDS.
KW Hypothetical protein.
FT DOMAIN 1 96 FRAME-1.
FT DOMAIN 100 194 FRAME-2.
FT DOMAIN 198 292 FRAME-3.
FT DOMAIN 296 391 FRAME-4.
FT DOMAIN 395 489 FRAME-5.
FT DOMAIN 493 587 FRAME-6.
SQ SEQUENCE 587 AA; 63703 MW; 3EAB3E3E3929203 CRC64;

Alignment Scores:
Pred. No.: 1 49e-33 Length: 587
Score: 410.50 Matches: 169
Percent Similarity: 29.74% Conservative: 35

Best Local Similarity: 24.64% Mismatches: 133
Query Match: 4.26% Indels: 350
DB: 1 Gaps: 14
US-09-966-880A-9 (1-5514) x ALU2_HUMAN (1-587)
QY 5509 AGGTCAGGAGTTTGTAGACGACGCTGCCCAACATGGTGAAACCCCTCTCTACTAAAAATA 5450
Db 219 ArgSerGlyAspArgAspHisProGly**HisGlyGluThrProSerLeuLeuLysIle 238
QY 5449 CAAACATTAGCCGAGTGTGGTGGCGCCCTGTATCCCATCTACCCGGGAGCGCTGAGG 5390
Db 239 GlnLysLeuAlaGlyArgGlyGlyArgLeu***SerGlnLeuLeuGlyArgLeuArg 258
QY 5389 CGACAGCTTTGCTTGAACCCAGGAGTGGAGGTTGCAGTGCAGTATGRTCATGCCATTC 5330
Db 259 GlnGluasnGlyValasnProGlyGlyGlyAlaCysSerGluProArgSerArgHisCys 278
QY 5329 ACTCCAGCATGGGCAACAAAGACCAATCCCATCTCTCAAAAAAATAAAAAA 5270
Db 279 ThrProAlaTriaAlaThr-GluArgAspSerValSerLysLys----- 292
QY 5269 AGAAGAGTAATACTTGTATCATCTGCAAACTTGAATACTATGCAAAAAATAGAAGACTCT 5210
Db 292 ----- 292
QY 5209 AAATTTTAAATATTAAAGATTTTAAATATCTCAAAATATCATAGTCTCTAATAATAATA 5150
Db 292 ----- 292
QY 5149 TTAGACTTAGTGTCTAATGACAATGACTGTGAAGATTCTTGTGTTTATTCTCTAGTCTT 5090
Db 292 ----- 292
QY 5089 TCTACAGGCACATATTTACATGATTTGGATTATATGATGATGATGATGATGATGATGATG 5030
Db 293 ----- 293
QY 5029 TGCTGCTTTTAACTAGTCAAAATCATCTAAATATCTTGTATGTTGTTCCACATAAT 4970
Db 294 ***PhePheLeuArgArgSerLeuAlaLeuSerProArgLeuGluCysSerGlyAlaIle 314
QY 4969 GACTGATTTCT 4910
Db 314 eSerAlaHisCysLysLeuArgLeuProGlySerArgHisSer----- 328
QY 4909 CACTAATTTGAATAGCATTTGTGGAATATCTCTGGTAACCCCAAGTTTCTCTTAAAGAAATC 4850
Db 328 ----- 328
QY 4849 AAGAGTCCCATTTGCAGCTGCAGACAGCAACACAGCTGTACAGCGTGGTTGATGATAA 4790
Db 328 ----- 328
QY 4789 GCACAAGTGATGCTGCGATTTTGTCCCTAGTCTTTTCCAGAAATGCATTTATCCAAATTG 4730
Db 328 ----- 328
QY 4729 TGGAGCTAGTGTGCACCTAATGGTACGGGGCAGAGAGTGCAGAGAAATTTGCCAAGTAA 4670
Db 329 ----- 329
QY 4669 GGGCTCAGGAATTTGAATAACGTTTGTGCGCAAGAGAAATTTGAAACTTTTCTTACACATTA 4610
Db 339 rGlyAla----- 341
QY 4609 TTTTCTTACATGAGGTATAATTATACATACAGTAATACTCCACCATTTTAAGATGCATA 4550
Db 341 ----- 341
QY 4549 ATTCTACGAATATTTTGGATGTAAACACCCCAAACTGCTACCACAGTACAGGATACAGAA 4490
Db 341 ----- 341

QY 4489 CCTGAATCACACCCCTCACCCACCAGAGCTTCTTCCCTCTTTGCGACTCTTCCCTCCC 4430
 Db |||||
 342 ---ArgHisAlaArg----- 346
 QY 4429 CACAGGACCACTGATTTGCTTTTTCCTTCCCAATTTTGTGGAGATGGAGTTTCA 4370
 Db |||||
 347 -----LeuIlePheValPheLeuValGluThrGlyPheH1 358
 QY 4369 CTGTGTTGCCAGGCTGTCTGAACCTCCTGAGTCTAGCAATCTTCCACGTCGCTCT 4310
 Db |||||
 358 sArgValSerGlnAspGlyLeuAspLeuThrSer***SerAlaArgLeu-GlyLeuP 378
 QY 4309 CAAAGTGTGGATTACAGGTGTGAGCCT----- 4281
 Db |||||
 378 roLysCysTrpAspTyrArgArgGluProProArgProAla*****PhePhe***A 398
 QY 4280 -----CTATGCTGCCCTGATTTGCTTCTTCTGTAATAAGTTACTTTCATTTT 4232
 Db |||||
 398 spGlyValSerLeuCysArgProGlyTrpSerAlaValAlaArgSerArgLeuThrAla 418
 QY 4231 CTAGAATTTCTGATGGCATATATCTCCCTTCTCATGCTCTGACTCTTTCATTAGCGT 4172
 Db |||||
 418 erSerAlaSer---ArgValHisAlaIleLeuLeu----- 428
 QY 4171 CATGACTTTGGGATTCATCCCAAGTAGCTTCTTCTCTT-----TGTACTGCGAC 4124
 Db |||||
 429 -----ProGlnProProGlu***LeuGlyLeuGlnAlaProAlaThrThrProG 445
 QY 4123 GGTGGTATTCATTGCGATGGATATAGCACATTTTGTATTGTTGTTCACTTGTGATGGAT 4064
 Db |||||
 445 ly***Phe----- 447
 QY 4063 ATTTGGTAGTCTTACTTTGGGGTTTGTGAATAAAGTACTATGACATTCACGTACC 4004
 Db |||||
 448 -----LeuTyrPhe*****Arg----- 453
 QY 4003 AGTCTTTGTGGATATACATGTAGAAATGTATTTAATAATAGTGCAGGTGACAGA 3944
 Db |||||
 454 -----ArgG 455
 QY 3943 CCTGACTTTAAAGTTAAATACCAAAATATACAGCTTAACTAAGTAGTCTATCATTAGGACA 3884
 Db |||||
 455 lyPheThrValLeuAlaArgMetValSerIleSer***ProArg-----Asp 471
 QY 3883 ACAGCTCTTGTAGTGTGTTCCATTTATTTTAAATTTTCTCTP----- 3840
 Db |||||
 471 ro-----ProAlaSerAlaSerGlnSerAlaGlyIleThrGlyVal 485
 QY 3839 -----TATTTTGTAGGCAAGGTCTGCTGCTCACCCAGGCTG 3800
 Db |||||
 485 erHisArgAlaArg*****PhePheGluThrGluSerArgSerValAlaGlnAlaG 505
 QY 3799 GAGTGCAGTGGCGAAGCACTGCTCAGTGCAGCCTTGACCTCTGGGCTCCAAAGATTCT 3740
 Db |||||
 505 lyValGlnTrpArgAspLeuGlySerLeuGlnAlaProProGlyPheThrProPheS 525
 QY 3739 CCTGCTCAGCCC-CCTAGTAGCTGGGACACAGCACATGCCACCATGCCCGCTAATT 3681
 Db |||||
 525 erCysLeuSerLeuProSerTrpAspTyrArgArgProProArgProAlaAsn 545
 QY 3680 TTTGTTAT-----AGAAACAGGTTTGGCCATTTGCTCAGGCTGTTGTAAGTCCAG 3627
 Db |||||
 545 heCysIlePheSerArgAspGlyValSerProCys***ProGlyTrpSerArgSerProA 565
 QY 3626 GGCTCAAGCATCACCCACCTCAGCCTCCAGAGTGTGGGATATACAGGCGTGAGCCAC 3567
 Db |||||
 565 spLeu--ValIleArgProArgProProLysValLeuGlyLeuGlnAla***AlaTh 584
 QY 3566 CACACCTGGC 3557
 Db |||||
 584 rAlaProGly 587

RESULT 12
 ALU3_HUMAN
 ID ALU3_HUMAN STANDARD; PRT; 587 AA.
 AC P39190;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alu subfamily SBI sequence contamination warning entry.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95021758; PubMed=7935834;
 RA Claverie J.-M., Makalowski W.;
 RT "Alu alert."
 RL Nature 371:752-752(1994).
 RN [2]
 RP CONCEPT.
 RX MEDLINE=92241891; PubMed=1572661;
 RA Claverie J.-M.;
 RT "Identifying coding exons by similarity search: alu-derived and other potentially misleading protein sequences.";
 RL Genomics 12:838-841(1992).
 RN [3]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=88333009; PubMed=3138422;
 RA Quentin Y.;
 RT "The Alu family developed through successive waves of fixation closely connected with primate lineage history.";
 RL J. Mol. Evol. 27:194-202(1988).
 RN [4]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=91178815; PubMed=1706781;
 RA Jurka J., Milosavljevic A.;
 RT "Reconstruction and analysis of human Alu genes.";
 RL J. Mol. Evol. 32:105-121(1991).
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.
 CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON. 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES. CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO ACID SEQUENCES.
 CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER, CONA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CONAS LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS, CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES BEING REPORTED.
 CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE CODING NUCLEOTIDE SEQUENCE.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

QY 5175 TTGAGATTAAAAAATCTTTAATAATTTTAAATTTTAGAGCTCTTCTATTCTTCATAGTA 5234
Db 437 ----- 437
QY 5235 TTCAAGTTTGACAAATGATCAAGTATTAATCTTCTTTCTTTTCTTTTCTTTTCTTTT 5284
Db 438 -----GlnAlaProAlaThrThrProGly***PheLeuTy*Phe*****ArgA 454
QY 5284 ----- 5284
Db 454 rgGlyPheThrValLeuAlaGlyMetValSerIleSer***ProArgAspProProAlas 474
QY 5285 -----TTTCTTTT 5291
Db 474 erAlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArg*****PheP 494
QY 5292 TTGAGATGGAGTTTGGTCTTGTGGCCATCTCGATGGATGGATGGATGGATGGATGGAT 5351
Db 494 heGluThrGluSerArgSer-ValAlaGlnAlaGlyValGlnTrpArgAspLeuGlySer 513
QY 5352 CTGCAACTCTCCACTCTCGGTTCAGCAAGCTGTGCGCTCAGCTCCCGGGTAGATGG 5411
Db 514 LeuGlnAlaProProGlyPheThrProPheSerCysLeuSerLeuProSerSerTrp 533
QY 5412 GATTACAGGCGCCACCACACACTCGGCTAATGTTTGTATTCTTTAGTAGAGATGGGTT 5471
Db 534 AspyrArgArgProProLeuArgProAlaAsnPheCysIlePheSerArgAspGlyVal 553
QY 5472 TCACCATGTTGGCAGGCTGTCTCAAACTCTGACCTC 5510
Db 554 SerProPhe***ProGlyTrpSerArgSerProAspLeu 566

RESULT 13

ALUS_HUMAN STANDARD; PRT; 585 AA.
ID ALUS_HUMAN STANDARD; PRT; 585 AA.
AC P39192;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SC sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert."
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RL closely connected with primate lineage history."
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes."
RL J. Mol. Evol. 32:105-121(1991).
CC -I- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX

CC CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -I- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -I- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -I- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -I- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14571; -; NOT_ANNOTATED_CDS.
DR KW Hypothetical protein.
FT DOMAIN 1 95 FRAME-1.
FT DOMAIN 99 193 FRAME-2.
FT DOMAIN 197 291 FRAME-3.
FT DOMAIN 295 389 FRAME-4.
FT DOMAIN 393 487 FRAME-5.
FT DOMAIN 491 585 FRAME-6.
SQ SEQUENCE 585 AA; 63957 MW; 46EE8C4F493650A7 CRC64;

Alignment Scores:
Pred. No.: 4.78e-30 Length: 585
Score: 376.50 Matches: 176
Percent Similarity: 29.39% Conservative: 25
Best Local Similarity: 25.73% Mismatches: 130
Query Match: 3.91% Indels: 354
DB: 1 Gaps: 13

US-09-966-880A-9 (1-5514) x ALUS_HUMAN (1-585)
QY 5494 GACCAGCTGCCAACATGGTGAACACCCATCTCTACTATAAATACAAACATTACCGAG 5435
Db 223 AsphisProGlyGlnHisGlyThrProSerLeuLeuLysIleGlnLysLeuAlaGly 242
QY 5434 TGTGGTGGTGGCGCTGTAATCCATCTACCGGAGGCTGAGCGACAGCTTTGCTTG 5375
Db 243 ArgGlyGlyAlaArgLeu***SerGlnLeuLeuGlyArgLeuArgGlnGlnAsnArgLeu 262
QY 5374 AACCCAGAGCTGGAGTTCAGTCAGCTATGTCATGCCATTCACCTCCAGCATGGCA 5315
Db 263 AsnProGlyGlyGlyGlyCysSerGluProArgSerArgHisCysThrProAlaTrp--A 282
QY 5314 ACAAGACCAAACTCCATCTCAAAAAAATAAAAAAAAAAAAAAAAAAAGAGTAATACT 5255
Db 282 rgGlnSerGluThrProSerGlnLysLys*****----- 293
QY 5254 TGATCATTTGCAAACTTGAATACTATGAAAAAATAGAGAGCTCTAAATTTTAAATATT 5195
Db 293 ----- 293

QY	5194	AAAGATTTTAAATCTCAATATCATAGTGTCTAATAATAATAGACTTAGTGTCT	5135
Db	294	-----Phephe-----	296
QY	5134	AATGCAATGACTGTAGATTCTGTGTGTTATTCCTCTAGTCTTCTACAGGCACATAT	5075
Db	297	-----LeuArgArgSerLeuAlaLeu-----	303
QY	5074	TTACATGATTTGGATATACATTCATGCTGTGATTTTGTACTCTGCTGCTTTTAAAC	5015
Db	303	-----	303
QY	5014	TAGTCAAAATGATCAATATATCTTGCTATGTGTTGCCACATAAATGACTGATTTCTCCTC	4955
Db	304	-----SerProg	306
QY	4954	CTCTTTGTTTCTGGCTCTACGACAGTCTCTTGCATATCATAAAGCACTAAATTTGAATAG	4895
Db	306	lyTrpSerAlaValAlaArgSerArgLeu-----	315
QY	4894	CATTTGTGGAATACTCTGTGTAACCAAGTTTCCTTAAAGAAATCAAGAGTCCCATTCG	4835
Db	315	-----	315
QY	4834	AGCTGCAGACAGCAACAGCTGTACAGCGTGTGTATATGAAAGCAACAAGTGCCT	4775
Db	316	-----ThrAlaThrSerAla-----	320
QY	4774	GGATTTGTCCTAGTCTTTTCCAGAAATGCATTTATCCAAATTTGTGGAGCTAGTCTGT	4715
Db	321	-----SerArg-----ValGlnAlaIleLeuLeuProGlnP	331
QY	4714	CACCTAATGTTAGGGGCAAGAGTGCAGAGAAATTTGCGAAGTAAAGGCTCAGGAATTTG	4655
Db	331	roProGlu***LeuGlyLeuGlnAlaArgAlaThrThrProSer***PheLeuTyPhe*	351
QY	4654	AATAAGCT-----TTGCTGGCAAGAGAATTTGAACACTTTTCTTACACATTTATTT	4607
Db	351	****ArgArgGlyPheThrMetLeuAlaArgMetValSerIle-----	365
QY	4606	TTTCTTACATTTGAGGTATATTTACATACAGTAAACTCCACCATTAAAGTGCATAATT	4547
Db	366	-----Ser***ProArgAspProProAlaSerAlaSerGlnS	378
QY	4546	CTACGAATATTTTGGATGTAAACAC-----	4520
Db	378	erAlaGlyIleThrGlyValSerHisArgAlaArg*****Phephe***AspGlyV	398
QY	4519	CCAACTGTTACACAGTCAGGATACAGAACCTGAATCACCACCTCACCCCCACAGAGC	4460
Db	398	alSerLeuGlyValGlnAlaGlyValGlnTrp-----ArgA	410
QY	4459	TTCTTCCCTTTGCACTCTTCCCTCCCTCCACAGGACACCACTGATTTTGCTTTCTTTT	4400
Db	410	spleuGlySerLeuGln-ProProProProGlyPheLysArgPheSerCysLeuSerLeu	429
QY	4399	-----TTTCCCATTTTGTG	4385
Db	430	ProSerSerTrpAspTyArgArgAlaProProArgProAlaAsnPheCysIlePheSer	449
QY	4384	GAGAATGGAGTTTCACTGTGTTCCAGCGTGTGCTTGAACCTCCTGAGTTCTAGCAATCT	4325
Db	450	ArgAspGlyValSerProCysTrpProGlyTrpSer-ArgSerLeuAspLeuValIleAr	469
QY	4324	TCCCAGCTGGCCCTCAAAAGTGTGGATTACAGGTGTGAGCCTCTATGCTGCCCTGAT	4265
Db	469	gProProArgProProLysValLeuGlyLeuGlnAla***AlaThrAlaProGly-----	487
QY	4264	TTGCTTTTCTGTAATAAGTTACTTTTCATTTTCTAGAATTTTCGTATGGCATATATC	4205
Db	487	-----	487
QY	4204	TCCTCTTCATGTCTGCACTCTTTTCATTTAGCGTCATGACTTTTGGGATTTTCATCCCAAGTA	4145
Db	487	-----	487
QY	4144	GCTTCTTCTCTTACTGCAGGGTGGTATTTCCATTCGATGATATAGCACATTTTGTGTTA	4085
Db	487	-----	487
QY	4084	TTTGTTCACCTTGTGTGATGATATTTGGATAGTCTCTACTTTGGGGTTTGTGTAATAAGG	4025
Db	487	-----	487
QY	4024	TACTATGAACATTCACGTACCAGTCTTTGTGTGGATATACATTTGACAATTTGATATTTAA	3965
Db	487	-----	487
QY	3964	TAATGTAGTGCAGGTGCAGACCCCTGACTTTTAAAGTTTAAATACCAAAATATCAGCCTAA	3905
Db	487	-----	487
QY	3904	CTAAGTAGCTCATCTAGGACACAGCTCTTTGTAGTTGTTCATTTATTTTAAATTTT	3845
Db	488	-----	488
QY	3844	CTTCTTATTTTGTAGGCAAGTCTGCTGCTCCTCACCAGGCTGAGTGCAGTGGCGCA	3785
Db	488	*****PhePheGluThrGluSerArgSerValAla-ArgLeuGluCysSerGlyAla	507
QY	3784	AGCACTGCTCAGTGCAGCCTTGACCTCTGGGCTCCAAAGATTTCTCTGCTCAGGCCCC	3725
Db	508	IleSerAlaHisCysAsnLeuArgLeuProGlySerSerAspSerProAlaSerAlaSer	527
QY	3724	TA-GTAGCTGGGACCAGCAGCACATGCCACCATGCCCGGCTAAATTTTGT-----ATA	3672
Db	528	ArgValalaglyThrThrGlyAlaArgHisHisAlaGlnLeuIlePheValPheLeuVal	547
QY	3671	GAACAGGGTTTTCATGTTGCTCAGCTGGTCTTTGAACCTCCAGGGCTCAACGGATCCA	3612
Db	548	GluThrGlyPheHisHisValGlyGlnAspGlyLeuAspLeuLeuThrSer****-SerA	567
QY	3611	CCACCTCAGCCTCCACAGAGTGTGGGATTACAGGGGTGAGCCACACACACCTGGCT	3556
Db	567	laArgLeuGlyLeuProLysCysTrpAspTyArgArgGluProProArgProAla	585
RESULT 14			
ALU3_HUMAN STANDARD; PRT: 587 AA.			
ID	ALU3_HUMAN	STANDARD;	PRT: 587 AA.
AC	P39150;		
DT	01-FEB-1995 (Rel. 31, Created)		
DT	01-FEB-1995 (Rel. 31, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Alu subfamily SBI sequence contamination warning entry.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Eutheria; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID:9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=95021758; PubMed=7935834;		
RA	Claverie J.-M., Makalowski W.;		
RT	"Alu alert.";		
RL	Nature 371:752-752(1994).		
RN	[2]		
RP	CONCEPT.		
RX	MEDLINE=92241891; PubMed=1572661;		
RA	Claverie J.-M.;		
RT	"Identifying coding exons by similarity search: alu-derived and other		
RT	potentially misleading protein sequences.";		
RL	Genomics 12:838-841(1992).		
RN	[3]		
RP	ALU FAMILIES CLASSIFICATION.		
RX	MEDLINE=88333009; PubMed=3138422;		
RA	Quentin Y.;		

RT "The Alu family developed through successive waves of fixation
closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]

RP ALU FAMILIES CLASSIFICATION.

RX MEDLINE=91178815; PubMed=1706781;

RA Jurka J., Milosavljevic A.;

RT "Reconstruction and analysis of human Alu genes.";

RL J. Mol. Evol. 32:105-121(1991).

CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
REPEATS.

CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.

CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
ACID SEQUENCES.

CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS.

CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
BEING REPORTED.

CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CODING NUCLEOTIDE SEQUENCE.

CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----

CC EMBL: U14569; -- NOT_ANNOTATED_CDS.

KW Hypothetical protein.

FT DOMAIN 1 96 FRAME-1.

FT DOMAIN 100 194 FRAME-2.

FT DOMAIN 198 292 FRAME-3.

FT DOMAIN 296 391 FRAME-4.

FT DOMAIN 395 489 FRAME-5.

FT DOMAIN 493 587 FRAME-6.

SQ SEQUENCE 587 AA; 63573 MW; 85C4155726DEF235 CRC64;

Alignment Scores:

Pred. No.: 1.33e-28 Length: 587
Score: 362.50 Matches: 167
Percent Similarity: 29.82% Conservative: 37
Best Local Similarity: 24.42% Mismatches: 136
Query Match: 3.77% Indels: 346
DB: 1 Gaps: 15

US-09-966-880A-9 (1-5514) x ALU3_HUMAN (1-587)

Qy 5509 AGGTCAGAGTTTGAGACCGCTGGCCCAACATGGTGAACCCCATCTCTACTAAAAATA 5450
||||| ||| ||||| ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 219 ArgSerGlyAspArgAspHisProGly***AsnGlyGluThrProSerLeuLeuLysIle 238
||||| ||| ||||| ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 5449 CAAACATATAGCCAGTGTGGTGGCGCTGTATCCCATCTACCCGGGAGGCTGAGG 5390
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 239 GlnLysLeuAlaGlyArgSerGlyArgLeu***SerGlnLeuLeuGlyArgLeuArg 258

Qy 5389 CGACAGCTTTGCTTGAACCCAGGAGGTGGAGTTGTCAGTGTATGTCATGCCATCC 5330
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 259 GlnGluAsnGlyValAsnProGlyGlyGlyAlaCysSerGluProArgSerArgHisCys 278
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 5329 ACTCCAGCATGGGCAACAGACCAAACTCCATCTCAAAAACAAAAACAAAAACAAAA 5270
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 279 ThrProAlaTrpAlaThr-GluArgAspSerValSerLysLys ----- 292
ThrProAlaTrpAlaThr-GluArgAspSerValSerLysLys ----- 292

Qy 5269 AGAAAGAGTAATACTTGCATCTGTCAAACTTGAATACTATATGGAATAATAAGAGCTCT 5210
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 292 ----- 292
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 5209 AAATTTTAAATATTAAGATTTTTTAAATCTCAAAATATCATAGTGTCTAATAATAATA 5150
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 292 ----- 292
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 5149 TTAGACTTAGTGTCTAATGACATGACTGTTAAGATTCTTGTGTGTTTATTCCTCTAGTCTT 5090
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 292 ----- 292
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 5089 TCTACAGGCACATATTTACATGATTGATTATATCTGATTCATGTCATGTGTGATTTGTACTC 5030
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 293 ----- 294
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 5029 TGCTGCTTTTAACTAGTCAAAATGATCTAAATATCTTGTCTATGTTGTTCCACATAAT 4970
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 294 ***PheLeuArgArgSerLeuAlaLeuSerProArgLeuGluCysSerGlyGlyI 314
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 4969 GACTGATTTCTCCCTCTCTTTGTTTCTGGCTCTAGCACAGTCTCTTGCATATCATTAAG 4910
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 314 eSerAlaHisCysLysLeuArgLeuProGlySerArgHisSer ----- 328
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 4909 CACTAATTTGAATAGCATTTGTGGAATACTCTGTAACCAACCAAGTTTCCTTAAAGAAATC 4850
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 328 ----- 328
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 4849 AAGAGTCCCATTCGAGCTGCAGACAGCAACACAAGCTGTACAGCGTGGTTGATATGAAA 4790
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 329 -----ProAlaSerAlaSer ----- 333
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 4789 GCACAAGTGATGCGCTGGATTTTGTCCCTAGTCTTTTCCAGAAATGCAATTTATCCANTTG 4730
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 333 ----- 333
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 4729 TGGAGTAGTGTCTCACCTAATGTTGCGGGCAGAAAGTGCAGAGAAATTTCCGGAAGTAAA 4670
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 334 -----GlnValAl 336
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 4669 GGGCTCAGGAATTTGAATAACGTTTGTGCGCAAGAGAATTGAAACTTTTCTTACACATTA 4610
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 336 aGlyThr -----ThrGlyAlaArgHisTy 344
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 4609 TTTTCTTACATTGAGGTATTAATTTACATACAGTAACCTCCACCCATTTTAAGATGCATA 4550
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 344 r ----- 344
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 4549 ATTCTACGAATATTTTGGATGTAAACACCCCAAACTGCTACCACAGTCAGGATACAGAA 4490
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 344 ----- 344
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 4489 CCTGAATCACCCCTCCACCCACCAGAGCTTCTTCCCTCTTTCGAGTCTTCCCTCCCCTCC 4430
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 344 ----- 344
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 4429 CACAGGACCACTGATTTGCTTTTCCATTTTGTGGAGAATGGAGTTTCA 4370
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 345 -----AlaArgLeuIlePheValPheLeuValGluThrGlyPheHi 358
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 4369 CTGTGTTGCCAGGCTGTCTGAACCTCTAGTCTAGCAATCTTCCCACGTGCGCTCT 4310
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 358 sArgPheSerArgAspGlyLeuAspLeuLeuThrSer***SerAlaArgLeu-GlyLeuP 378
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

QY 4309 CAAAGTCCTGGATTACAGGTGTGACCT----- 4281
Db 378 roLysCysTrpAspTyrArgGluProProArgProAla*****PhePhe***A 398
QY 4280 -----CTATGCGCTGCCTGATTGCTTCTGTAACTATAAGTTACTATTCATTTT 4232
Db 398 spGlyValSerLeuCysArgPro-----GlyTrpSerAlaValAlaGlyS 413
QY 4231 CTAGAATTTCGTGGC-----ATACATTA-----TCTCTCCCTT 4197
Db 413 erArgLeuThrAlaSerSerAlaSerArgValHisAlaIleLeuLeuProGlnProProL 433
QY 4196 CATGCTGCTACTCTTCATTTACGTCATGACTTTGGGATTCATCCCAAGTAGCTTCCTC 4137
Db 433 ys***LeuGlyLeuGlnAlaProAlaThrProGly----- 445
QY 4136 CTTTGTACTGCAGGTGGTATTCCATTGTCATGATATAGACATTTGTTTATTG--- 4081
Db 446 -----***PheLeuTyrPhe*** 452
QY 4080 -----TTCACCTTGTGATGATATTTGGATAGTCTACTTTGGGCTTTGTGA 4032
Db 452 **ArgArgGlyPheThr-----ValL 459
QY 4031 ATAAAGTACTATGAACATTCACGACACAGCTCTTTGTGGATATACATTGAGAATTGAT 3972
Db 459 euAlaGlyMetValSerIleSer***Pro----- 468
QY 3971 ATTTTAATAATGATGTCAGGTCAGACAGCCCTGACTTTAAAGTTAAATACCAAAATATC 3912
Db 469 -----ArgAspProPro----- 472
QY 3911 AGCCTAACTAGTACTCATCTAGGACACAGCTCTTTGTAGTCTTTCCATTTATTTT 3852
Db 473 -----AlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHis-Arg 487
QY 3851 RAATTTTCTCTTATTTTGGCAAGGTCTGCTGTCACCCAGCGTCGAGTCGAG 3792
Db 498 AlaArg*****PhePheGlyThrGluSerArgSerValAlaGlnAlaGlyValGln 507
QY 3791 TGGCGCAAGCAGCTGCTAGTCAGCAGCTTGACCTCTGGGCTCCAAAGATTCTCCCTGCCTC 3732
Db 508 TrpAlaGspLeuGlySerLeuGlnAlaProProGlyPheThrProPheSerCysLeu 527
QY 3731 AGCCC-CTAGTAGTGGGACACAGGACATGCCACCATGCCCGCGTAATTTTGTAT 3673
Db 528 SerLeuProSerSerTrpAspTyrArgArgProProLeuArgProAlaAsnPheCysIle 547
QY 3672 -----AGAAACAGGTTTTGCCATGTTGCTCAGCTGTCTTGAACCTCCAGGGCTCAAG 3619
Db 548 PheSerArgAspGlyValSerProPhe***ProGlyTrpSerArgSerProAspLeu--V 567
QY 3618 CGATCCACCCACCTCAGCTCCAGAGTCTGGGATTACAGCGTGAGCCACACACCTG 3559
Db 567 alIleArgProProArgProProLysValLeuGlyLeuGlnAla***AlaThrAlaProG 587
QY 3558 GC 3557
Db 587 ly 587

RESULT 15
ALU4_HUMAN
ID ALU4_HUMAN STANDARD; PRT; 603 AA.
AC P39191.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SB2 sequence contamination warning entry.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]

```

Alignment Scores:

```

RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RN CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RX ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RX ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -!- CAUTION: ALU REPTITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LOCATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14570; -; NOT_ANNOTATED_CDS.
KW Hypothetical protein.
FT DOMAIN 1 98 FRAME-1.
FT DOMAIN 102 199 FRAME-2.
FT DOMAIN 202 300 FRAME-3.
FT DOMAIN 304 401 FRAME-4.
FT DOMAIN 405 502 FRAME-5.
FT DOMAIN 506 603 FRAME-6.
SQ SEQUENCE 603 AA; 65272 MW; BBAAD0AD46BEA114 CRC64;

```


Qy	4470	GGTGAGGGTGGTGATTCAGGTTCTGTATCTGACTGTGGTAGCAGTTTGGGGTGTTTACA	4529
Db	268	ArgGluProGlyLysArgSer	274
Qy	4530	TCCAAAAATATTTCGTAGAAATTTATGCATCTTAAATGGGTGGAGTTTACTGTATGTAAATTA	4589
Db	275	-----LeuGln**AlaGlu-----IleAlaProLeu	283
Qy	4590	TACCTCAATGTAAGAAAAATAATGTGTAAGAAAAATTTCAATTTCTCTTCCACAGCAACG	4649
Db	284	GlnSerAlaValargProGly-----ArgGlnSerGluThrProSerGlnLysLys	300
Qy	4650	TTATTCAAAATTCCTGAGCCCTTTACTTCGCAAAATTTCTCTGCACATTTGCGCCCGTACCATTT	4709
Db	301	*****Phe-----PheLeuArgArgSerLeuAlaLeuSerProArgPro---	316
Qy	4710	AGGTGACAGCAGTCTCCACAAATTTGGATAAATGCATTTTCTGGAAAAAGACTAGGGACAA	4769
Db	316	-----	316
Qy	4770	AATCCAGGCATCATTGTGCTTTTCATATCAACACGCTGTACAGCTTGTGTGCTGCTG	4829
Db	317	-----AspCysGly---	319
Qy	4830	CAGCTGCAATGGGAGCTCTTGATTTCTTTAAAGAAA-----	4865
Db	320	---LeuGlnTrpArgAsnLeuGlySerLeuGlnAlaProLeuProGlyPheThrProPhe	338
Qy	4866	-----CTTGGGTTTACCAGAGTATTTCCACAAATGCTATTCAAATTTAGTCTTATGATAT	4919
Db	339	SerCysLeuSerLeuProSerSerTrpAsp-----Tyr	349
Qy	4920	GCAAGCACCTGTCTAGGAGCCAGAAA-----ACAAAGAGGAGGAGGAAATTCAGTCA	4970
Db	350	ArgArgProProArgProAlaAsnPheLeuTyPhe*****ArgArgGlyPheThr	369
Qy	4971	TTATGTGGGAACACATPAGCAAGATATTTAGATCATTTTGACPTAGTTTAAAAAGCAGCAG	5030
Db	370	LeuLeuAlaArgMetValSerIleSer**ProHisAspProProAlaSerAlaSerGln	389
Qy	5031	AGTACAAAAATCACACATGCAATCAGTATATCCAAATCATGTAATATGTGCCTGTGTAGAA	5090
Db	390	SerAla-Gly-----	392
Qy	5091	AGACTAGAGGAATAAACACAGAATCTTAACAGTCATTGTTCATTAGACACTAAGTCTAAT	5150
Db	392	-----	392
Qy	5151	TATTATTATTAGACACTATGATATTTGAGATTTAAAAAATCTTTAATATTTTAAAAATTA	5210
Db	392	-----	392
Qy	5211	GAGCTCTCTATTTTCCATAGTATTTCAAGTTTGCACAAATCATCAAGTATTACTCTTTCTT	5270
Db	393	-----IleThrGlyValSerHisArgAlaArg**	402
Qy	5271	TTTTTTTTTTTTTTTTTTTTTTTTCAGATGGAGTTTGGTCTGTGTGGCC-----ATGCT	5324
Db	402	*****PhePhe*****AspGlyValSerLeuCysArgProGlyArgThrAl	419
Qy	5325	GGAGTGAATGGCATGAYCATAGCTACTGCAACCTCCACCTCCTCGGTGTCAAGCAAGC	5384
Db	419	AspCysSerGlyAlaIleSerAlaHisCysLysLeuArgPheProGlySerArgHisSe	439
Qy	5385	TGTCGCTCAGCTCCCGGTAGATGGGATTTACAGGGCCGCCACCACACACTCGGCTAAT	5444
Db	439	rProAlaSerAlaSerProValAlaGlyThrThrGlyAlaArgHisArgAlaArgLeuIl	459
Qy	5445	GTT--TGATTTTTTAGTAGAGATGGGGTTTCCACATGTTGGCCAGGCTGGTCTCAAACTCC	5503
Db	459	epheCysIlePheSerArgAspGlyValSerProCys**ProGlyTrpSerArgSerPr	479
Qy	5504	TGACCTC	5510

Db |||||||
 479 oAspleu 481

Search completed: June 14, 2003, 18:30:43
Job time : 136.164 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:12:03 ; Search time 338.102 Seconds
(without alignments)
6720.711 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 9659
Sequence: 1 acagcagaatacatgtcca.....tcaactcctgacctcagag 5514

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p_model -DEV=xlp
-Q=/cgn2_1/USPTO_spo3/US09966880/runat_14062003_175524_10316/app_query.fasta_1.9493
-DB=SPTRMBL_21 -QMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09966880.ecgn.1.1.978 @runat_14062003_175524_10316 -NCPU=6 -ICPU=3
-NO_WMAP -LARGESQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvrius:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
c 1	309.5	3.2	171	4 Q9H728	Q9H728 homo sapien

c	2	293.5	3.0	127	4	Q9H743	Q9H743
c 3	290	3.0	152	4	Q9NX85	Q9NX85	homo sapien
c 4	289	3.0	118	4	Q9H387	Q9H387	homo sapien
c 5	287	3.0	115	6	Q9N083	Q9N083	macaca fasc
c 6	282	2.9	239	4	Q9NX17	Q9NX17	homo sapien
c 7	277	2.9	118	4	Q9H387	Q9H387	homo sapien
c 8	274	2.8	136	4	Q9N086	Q9N086	homo sapien
c 9	267.5	2.8	152	4	Q9NX85	Q9NX85	homo sapien
c 10	263	2.7	162	4	Q9BYA5	Q9BYA5	homo sapien
c 11	260	2.7	72	4	Q8TB48	Q8TB48	homo sapien
c 12	260	2.7	136	4	Q9N086	Q9N086	homo sapien
c 13	256.5	2.7	163	4	Q9GMM0	Q9GMM0	homo sapien
c 14	255.5	2.7	130	4	Q9HBS7	Q9HBS7	homo sapien
c 15	251	2.6	109	4	Q8WZ39	Q8WZ39	homo sapien
c 16	251	2.6	232	4	Q9H5R3	Q9H5R3	homo sapien
c 17	251	2.6	238	4	Q8WTZ3	Q8WTZ3	homo sapien
c 18	250.5	2.6	666	4	P78525	P78525	homo sapien
c 19	246	2.5	238	4	Q8WTZ3	Q8WTZ3	homo sapien
c 20	246	2.5	239	4	Q9NX17	Q9NX17	homo sapien
c 21	245	2.5	66	4	Q96HL9	Q96HL9	homo sapien
c 22	245	2.5	169	4	Q9H397	Q9H397	homo sapien
c 23	245	2.5	208	4	Q9NW14	Q9NW14	homo sapien
c 24	244.5	2.5	120	4	Q9H6G8	Q9H6G8	homo sapien
c 25	244	2.5	122	6	Q9BGW3	Q9BGW3	macaca fasc
c 26	242.5	2.5	118	4	Q9P195	Q9P195	homo sapien
c 27	241.5	2.5	104	4	Q9P0E3	Q9P0E3	homo sapien
c 28	240.5	2.5	139	4	Q9BVD9	Q9BVD9	homo sapien
c 29	239.5	2.5	123	4	Q9HAD8	Q9HAD8	homo sapien
c 30	239	2.5	84	4	Q9UHT1	Q9UHT1	homo sapien
c 31	239	2.5	151	4	Q9HA67	Q9HA67	homo sapien
c 32	238	2.5	133	4	Q96JF5	Q96JF5	homo sapien
c 33	237	2.5	375	4	O60448	O60448	homo sapien
c 34	236	2.4	171	4	Q9H728	Q9H728	homo sapien
c 35	236	2.4	375	4	O60448	O60448	homo sapien
c 36	235	2.4	129	4	Q9H9H0	Q9H9H0	homo sapien
c 37	233	2.4	83	4	Q96ID7	Q96ID7	homo sapien
c 38	231	2.4	62	4	Q96FS0	Q96FS0	homo sapien
c 39	231	2.4	121	4	Q96N97	Q96N97	homo sapien
c 40	231	2.4	179	4	Q96MD7	Q96MD7	homo sapien
c 41	230	2.4	535	4	Q96EB1	Q96EB1	homo sapien
c 42	229.5	2.4	123	4	Q9PIN7	Q9PIN7	homo sapien
c 43	229	2.4	666	4	P78525	P78525	homo sapien
c 44	227.5	2.4	139	6	Q9BE58	Q9BE58	macaca fasc
c 45	227	2.4	687	4	Q9NXE7	Q9NXE7	homo sapien

ALIGNMENTS

RESULT 1
Q9H728 PRELIMINARY; PRT; 171 AA.
ID Q9H728
AC Q9H728;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE CDNA: FLJ21463 fis, clone COL04765.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okikani R., Oka T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK025116; BAB15071.1; -
SQ SEQUENCE 171 AA; 18436 MW; DDC546D275038FAF CRC64;

Alignment Scores: 1.19e-24 Length: 171
Pred. No.: 1

Score: 309.50 Matches: 74
 Percent Similarity: 64.06% Conservative: 8
 Best Local Similarity: 57.81% Mismatches: 33
 Query Match: 3.21% Indels: 14
 DB: 4 Gaps: 3

US-09-966-880A-9 (1-5514) x Q9H728 (1-171)

QY 3904 CTAAGTAGCTCATCTAGGACACAGCTCTTTGTAGTTGTTTCCATTTATTTTAAATTTT 3845
 Db 24 LeuLeuMetSerGlnLysLysLeuGlyPheCysGlyAsnPhe--LeuPheLeuAsnLeu 42
 QY 3844 CTTCTTATT-----TTTTTGAGGCAAGGTCGTGGTC 3815
 Db 43 AlaIleLeuGlnThrLysIleSerSerPhePhePheLeuArgGlnSerLeuThr 62
 QY 3814 CTGTCAACAGGCTGGAGTGGAGTGGCGCAAGCACTGTCACGTGCAGCTTGACCTCCTG 3755
 Db 63 LeuSerProArgLeuGluCysAsnGlyAlaIleSerAlaHiscysHisLeuArgLeuPro 82
 QY 3754 GGCTCCAAAGATTCTCTGCCCTCAGCCCCCTA-GTAGCTGGGACACAGGCACATGCCAC 3696
 Db 83 AspSerSerAsnSerProAlaSerAlaSerGlnValThrGlyIleThrGlySerHisHis 102
 QY 3695 CATGCCCGCGTAAATTTTGT-----ATAGAAACAGGTTTGGCATGTTGCTCAGGCT 3642
 Db 103 HisAlaTrpLeuIlePheValPheLeuValGluThrGlyPheCysHisValGlyGlnAsp 122
 QY 3641 GGTCTTGAACCTCCAGGCTCAAGGCTCAAGCCACTCAGCTGCCAGAGTCTGGGATT 3582
 Db 123 GlyLeuGluLeuLeuThrSerGlyAspProAlaSerAlaSerGlnSerAlaGlyIle 142
 QY 3581 ACAGCGCTGAGCCACCACTGG 3558
 Db 143 ThrGlyMetSerHisHisThrTrp 150
 RESULT 2
 Q9H743 PRELIMINARY; PRT; 127 AA.
 ID Q9H743
 AC Q9H743;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DE CDNA: FLJ21394 f1s, clone COL03536.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=COLON;
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
 RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK025047; BAB15056.1; -;
 SQ SEQUENCE 127 AA; 13754 MW; 25851EDBD01EF78 CRC64;

Alignment Scores:
 Pred. No.: 6 53e-23 Length: 127
 Score: 293.50 Matches: 64
 Percent Similarity: 74.23% Conservative: 8
 Best Local Similarity: 65.98% Mismatches: 20
 Query Match: 3.05% Indels: 5
 DB: 4 Gaps: 2

US-09-966-880A-9 (1-5514) x Q9H743 (1-127)

QY 3835 TTTTGGAGCAAGTCTGTCTGTACCCAGGCTGGAGTGGCGGCAAGCACTGCT 3776
 Db 7 PheLeuGlyGlnGlyLeuThrLeuSerProArgLeuGluCysSerSerThrIleSerAla 26

QY 3775 CAGTCAGCGCTTGACCTCTCTGGCTCCAAAGATTCTCCT---CCCTCAGCCCCCTAGTAG 3719
 Db 27 HisCysAsnLeuHisLeuLeuGlySerSerAsnSerProValAlaAlaSerPro--ValA 46
 QY 3718 CTGGGACACAGGCACATGCCACCATGCCCGCTAATTTTGT-----ATAGAAACAG 3665
 Db 46 laGlyThrThrGlyThrCysHisHisAspTrpLeuIlePheValPheLeuValGluThrG 66
 QY 3664 GGTTTGCCATGTTGCTCAGGCTGCTTGAACCTCCAGGCTCCAGCGCTCAAGCATCCACCT 3605
 Db 66 lypheHisHisIleGlyGlnThrGlyLeuGluPheLeuThrSerGlyAspProProThrL 86
 QY 3604 CAGCTCCAGAGTGTGGATTACAGCGTGTAGCCACCACTGG 3558
 Db 86 euAlaSerLysSerAlaGlyIleThrGlyValSerHisCysAlaTrp 101
 RESULT 3
 Q9NX85 PRELIMINARY; PRT; 152 AA.
 ID Q9NX85
 AC Q9NX85;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DE KAI0536 protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=ILEAL MUCOSA;
 RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
 RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
 RA Nakamura Y., Isogai T., Sugano S.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK000385; BAA91131.1; -;
 SQ SEQUENCE 152 AA; 16568 MW; 59065F45AAA301B5 CRC64;

Alignment Scores:
 Pred. No.: 1 66e-22 Length: 152
 Score: 290.00 Matches: 60
 Percent Similarity: 73.56% Conservative: 4
 Best Local Similarity: 68.97% Mismatches: 22
 Query Match: 3.00% Indels: 1
 DB: 4 Gaps: 0

US-09-966-880A-9 (1-5514) x Q9NX85 (1-152)

QY 5253 CAAGTATTACTCTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 5312
 Db 14 GlnValLeuPhePheLeuSerPheLeuPhePhePheLeuArgGlnSerPheAlaLeu 33
 QY 5313 GTTGCCCATCTGGAGTGGAGTGGCATGAYCATAGCTCACTGCAACTCCACCTCTCTGG 5372
 Db 34 ValAlaGlnAlaGlyValGlnTrpArgAsnLeuGlySerLeuGlnProProProGly 53
 QY 5373 TTCAAGCAAGCTGTCTGCCTCAGCTCCCGGTAGATGGATTACAGCGCCGCCACCA 5432
 Db 54 PheLysGlnPheSerCysLeuSerLeuSerSerTrpAspTrpArgHisAlaProPro 73
 QY 5433 CACTCGGCTAATGTTTGTATTTTCTAGATAGATGGGTTCACCATGTGGCCAGGCTGG 5492
 Db 74 CysProAla-TyrPheValPheLeuValAspMetGlyPheProHisValGlyGlnThrG 93
 QY 5493 TCTCAACTCTGACCTCA 5511
 Db 93 yLeuGluLeuLeuThrSer 99
 RESULT 4
 Q9H387 PRELIMINARY; PRT; 118 AA.
 ID Q9H387
 AC Q9H387;


```

QY 3745 GATTCTCTGCTCAGCCCCCTA-GTAGCTGGGACCACAGGCACATGCCACCATGCCCGG 3687.
Db :::::::::::::: ||| |||||||| ||||| |||||||| |||||||
149 AnSerProAlaLeuAlaSerGlnValAlaGlyIleThrGlyIleCysHisAlaArg 168
QY 3686 CTAATTTTGT-ATAGAAACAGGTTTGGCATGTGCTCAGGCTGCTGTGA 3633
Db :::::::::::::: ||| |||||||| ||||| |||||||| |||||||
169 GlnIlePheValGluThrGlyPheCysHisValGlyGlnAlaGlyLeuGlu 188
QY 3632 CTCAGGGCTCAAGCATCCACCCTCAGCTCCAGAGTCTGGGATTACAGCGTG 3573
Db :::::::::::::: ||| |||||||| ||||| |||||||| |||||||
189 LeuLeuIleSerGlyAspSerProAlaSerAlaPheGlnSerAlaGlyIleIleGlyVal 208
QY 3572 AGCCAC 3567
Db :::::::::::::: ||| |||||||| ||||| |||||||| |||||||
209 SerHis 210

RESULT 7
Q9H387 PRELIMINARY; PRT; 118 AA.
AC Q9H387:
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE PRO2550.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Zhang C., Yu Y., Zhang S., Wei H., Bi J., Zhou G., Dong C., Zai Y.,
RA Xu W., Gao F., Liu M., He F.;
RT "Functional prediction of the coding sequences of 75 new genes deduced
RT by analysis of cDNA clones from human fetal liver.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF130089; AAG35515.1; -
SQ SEQUENCE 118 AA; 13257 MW; 94688870CAC8760D CRC64;

Alignment Scores:
Pred. No.: 4.29e-21 Length: 118
Score: 277.00 Matches: 56
Percent Similarity: 74.68% Conservative: 3
Best Local Similarity: 70.89% Mismatches: 19
Query Match: 2.87% Indels: 1
DB: 4 Gaps: 0

US-09-966-880A-9 (1-5514) x Q9H387 (1-118)

QY 5277 TTTTNTTTTTTTTTTTCAGATGGAGTTTGGTCTTGTGTCCTGTCCTGAGTGGGAATGG 5336
Db ::::::::::||| |||||||| |||||||| |||||||| |||||||| |||||||| ||||||||
5 PhePhePheLeuPheLeuArgTrpSerPheThrLeuValAlaGlnAlaGlyValGlnTrp 24
QY 5337 CATGAYCATAGTCTACTCACTCCACCTCCTCGTTCACAGCAAGCTGTCGCTCAGC 5396
Db ::::::::::||| |||||||| |||||||| |||||||| |||||||| |||||||| ||||||||
25 ArgAspLeuSerSerProGlnProProProArgPheLeuArgPheSerCysLeuSer 44
QY 5397 CTCGGGGTAGATGGATTACAGCGCCACACACACTCGGCTTAATGTTGTATTTT 5456
Db ::::::::::||| |||||||| |||||||| |||||||| |||||||| |||||||| ||||||||
45 ProProSerSerTrpAspTrpArgHisAlaProProHisProAlaAsn-PheValPheLe 64
QY 5457 AGTAGATGGGTTTACCATTGTGGCCAGGCTGCTCAACTCCCTGACCTCA 5511
Db ::::::::::||| |||||||| |||||||| |||||||| |||||||| |||||||| ||||||||
64 uValGluThrGlyPheLeuArgValGlyGlnAlaGlyLeuGluLeuThrSer 82

RESULT 8
Q96NR6 PRELIMINARY; PRT; 136 AA.
AC Q96NR6:
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CDNA FLJ30278 fis, clone BRACE2002755.

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CEREBELLUM;
RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Kikuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K.,
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,
RT "NEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK054840; BAB70813.1; -
SQ SEQUENCE 136 AA; 14980 MW; D2336B649A110163 CRC64;

Alignment Scores:
Pred. No.: 9.5e-21 Length: 136
Score: 274.00 Matches: 54
Percent Similarity: 69.77% Conservative: 6
Best Local Similarity: 62.79% Mismatches: 25
Query Match: 2.85% Indels: 1
DB: 4 Gaps: 0

US-09-966-880A-9 (1-5514) x Q96NR6 (1-136)

QY 5510 GAGTCCAGGAGTTTGAGACCAGCTGGCCAAACATGGTGAACCCCATCTCTACTAAAAAT 5451
Db ::::::::::||| |||||||| |||||||| |||||||| |||||||| |||||||| ||||||||
52 GluValArgSerArgProAlaTrpProThrTrpArgAsnProIleSerThrLysAsn 71
QY 5450 ACAAAACATTAGCCGAGTGTGGTGGCGCTGTAATCCCATCTACCCGGAGGCTGAG 5391
Db ::::::::::||| |||||||| |||||||| |||||||| |||||||| |||||||| ||||||||
72 ThrLysIleSerGlnThrTrpArgMetProValValLeuAlaThrTrpGluAlaGlu 91
QY 5390 GCGACAGCTTTGCTTGAACCCAGAGGTGGAGTTCAGCTATGCTATGCTATTC 5331
Db ::::::::::||| |||||||| |||||||| |||||||| |||||||| |||||||| ||||||||
92 AlaGlyGlu-SerValAsnLeuGlyGlyArgGlyCysSerGluLeuArgArgCysArgC 111
QY 5330 CACTCCAGCATGGGCAACAGACCAAACTCCATCTCAAAAAAANAAAAA 5271
Db ::::::::::||| |||||||| |||||||| |||||||| |||||||| |||||||| ||||||||
111 sthrProAlaTrpAlaThrArgAlaLysLeuHisLeuArgLysAsnLysAsn 131
QY 5270 AAGAAAGAGTAACTACT 5255
Db ::::::::::||| |||
131 nGlyValThrLysThr 136

RESULT 9
Q9NX85 PRELIMINARY; PRT; 152 AA.
AC Q9NX85:
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE KIA0536 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ILEAL MUCOSA;
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Oka T.,
RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000385; BAA91131.1; -
SQ SEQUENCE 152 AA; 16568 MW; 59065F45AAA301B5 CRC64;

Alignment Scores:

```

Db 79 ValLeu-LeuTyrArgGlnAlaGlyValGlnTrpCysTyrLeuGlySerLeuGlnProLe 98

RP SEQUENCE FROM N.A.

1


```
Db      58 SerSerTyrSerProAlaSerAlaSerArgValThrGlyIleThrGlyMetCysHis 77
Qy      3692 GCCCGGTAATTTTGT- ----ATAGAAACAGGGTTTGGCCATGTTGCTCAGGCTGGT 3639
Db      78 AlaGlnLeuIlePheLeuValPheLeuValLysMetGluPheArgHisValGlyGlnThrSer 97
Qy      3638 CTTGAACCTCCAGGCTCAAGCGATCCACCCACCTCCAGCTCCAGAGTGGTGGATTACA 3579
Db      98 PheGluLeuLeuAlaSerSerProProAlaSerAlaSerGlnSerAlaGlyIleThr 117
Qy      3578 GCGGTGAGCCACCACACTGG 3558
Db      118 GlyValSerHisCysAlaTrp 124
```

RESULT 15

```
Q8WZ39
ID      Q8WZ39      PRELIMINARY;      PRT;      109 AA.
AC      Q8WZ39;
DT      01-MAR-2002 (TrEMBLrel. 20, Created)
DT      01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT      01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE      Hypothetical 12.3 kDa protein.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_taxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA      Wan D.F., Gu J.R.;
RT      "Novel human cDNA clones with function of inhibiting cancer cell
RT      growth.";
RL      Submitted (JUL-2000) to the EMBL/GenBank/DBDJ databases.
DR      EMBL; AF289553; AAL55737.1; -.
KW      Hypothetical protein.
SQ      SEQUENCE 109 AA; 12263 MW; 61560F7B74664B55 CRC64;
```

Alignment Scores:

Pred. No.:	3,16e-18	Length:	109
Score:	251.00	Matches:	54
Percent Similarity:	71.08%	Conservative:	5
Best Local Similarity:	65.06%	Mismatches:	23
Query Match:	2.60%	Indels:	1
DB:	4	Gaps:	0

US-09-966-880A-9 (1-5514) x Q8WZ39 (1-109)

```
Qy      5265 TTTCTTTT-----TTTTTTTGGATGGAGCTTTTGGTCTTGTGGCCATGCT 5324
Db      27 PheIleLeuValPhePheCysLeuPheLeuArgTrpSer-LeuAlaLeuLeuArgLe 46
Qy      5325 GGAGTGGAAATGGCATGAYCATAGCTACTGCAACCTCCACCTCTGGTTCAGCAAGC 5384
Db      46 uGluCysSerGlyAlaIleSerAlaHisCysAsnLeuHisLeuLeuGlySerTyrSe 66
Qy      5385 TGTGCGCTCAGCTCCCGGTAGATGGGATTACAGCGCCGCCACACACTCGCTAAT 5444
Db      66 rProValSerAlaSerGlnValAlaGlyThrGlyLeuCysHisAlaArgLeuAr 86
Qy      5445 GTTTGTATTTTATAGATAGATGGGGTTTCCACCATGTTGGCAGGCTGGTCTCAAACTGCT 5504
Db      86 gPheValPheLeuValGlnThrValPheHisHisValGlyGlnAlaGlyLeuLysLeuLe 106
Qy      5505 GACCTCA 5511
Db      106 uAlaSer 108
```

Search completed: June 14, 2003, 18:47:58
Job time : 353.102 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:13:03 ; Search time 53.1083 seconds
(without alignments)
6109.704 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 9659
Sequence: 1 acagacgaatacatagtgtcca.....tcaaaactctgacctcgag 5514

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_Spool/US0966880/runat_14062003_175525_10370/app_query.fasta_1.9493
-DB=Issued_Patents_AA -QMT=fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0966880 -CGN_1_1_139 -runat_14062003_175525_10370 -NCPU=6 -ICPU=3
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep: *
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep: *
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep: *
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep: *
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep: *
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	304	3.2	397	5 PCT-US95-17111A-121	Sequence 121, App
C 2	286	3.0	132	4 US-09-605-785-573	Sequence 573, App
C 3	255	2.6	132	4 US-09-605-785-573	Sequence 573, App
C 4	248.5	2.5	110	4 US-09-227-357-193	Sequence 168, App
C 5	238.5	2.5	99	4 US-09-288-143-168	Sequence 168, App
C 6	237	2.5	375	2 US-08-454-557C-121	Sequence 121, App
C 7	237	2.5	375	2 US-08-340-426D-121	Sequence 121, App
C 8	237	2.5	375	2 US-08-450-673C-121	Sequence 121, App
9	236	2.4	375	2 US-08-454-557C-121	Sequence 121, App
10	236	2.4	375	2 US-08-340-426D-121	Sequence 121, App
11	236	2.4	375	2 US-08-450-673C-121	Sequence 121, App
12	221	2.3	397	5 PCT-US95-17111A-121	Sequence 121, App

C 13	221	2.3	1079	3 US-09-058-489-22	Sequence 22, Appl
14	212	2.2	1079	3 US-09-058-489-22	Sequence 22, Appl
15	198	2.0	56	4 US-09-227-357-577	Sequence 577, App
16	192	2.0	99	4 US-09-288-143-168	Sequence 168, App
C 17	187	1.9	90	4 US-09-227-357-171	Sequence 171, App
C 18	181	1.9	110	4 US-09-227-357-193	Sequence 193, App
19	170	1.8	122	4 US-09-227-357-285	Sequence 285, App
C 20	169	1.8	122	4 US-09-227-357-285	Sequence 285, App
C 21	166.5	1.7	76	4 US-09-605-785-575	Sequence 575, App
22	156.5	1.6	76	4 US-09-605-785-575	Sequence 575, App
23	146	1.5	80	4 US-09-227-357-249	Sequence 249, App
C 24	139	1.4	72	4 US-09-227-357-655	Sequence 655, App
C 25	137.5	1.4	52	4 US-09-227-357-537	Sequence 537, App
C 26	136	1.4	388	4 US-09-265-630-11	Sequence 11, Appl
27	136	1.4	500	4 US-09-265-630-11	Sequence 11, Appl
C 28	128.5	1.3	80	4 US-09-227-357-249	Sequence 249, App
C 29	127	1.3	416	2 US-08-694-915-4	Sequence 4, Appl
C 30	124	1.3	35	4 US-09-288-143-170	Sequence 170, App
31	123.5	1.3	52	4 US-09-227-357-537	Sequence 537, App
C 32	123	1.3	47	4 US-09-227-357-656	Sequence 656, App
C 33	123	1.3	72	4 US-09-227-357-655	Sequence 655, App
C 34	117	1.2	56	4 US-09-227-357-577	Sequence 577, App
C 35	116	1.2	90	4 US-09-227-357-171	Sequence 171, App
C 36	113.5	1.2	70	4 US-09-188-930-131	Sequence 131, App
C 37	113	1.2	619	3 US-09-156-253-48	Sequence 48, Appl
C 38	113	1.2	619	5 PCT-US93-03027-6	Sequence 6, Appl
39	111.5	1.2	227	6 5498499-2	Patent No. 5498499
C 40	109	1.1	605	4 US-09-440-936-2	Sequence 2, Appl
C 41	108	1.1	500	4 US-09-265-630-13	Sequence 13, Appl
42	104	1.1	579	3 US-08-704-711A-1	Sequence 1, Appl
43	104	1.1	579	4 US-09-521-220-1	Sequence 1, Appl
44	103.5	1.1	43	4 US-09-461-697-127	Sequence 127, App
45	103	1.1	65	4 US-09-227-357-538	Sequence 538, App

ALIGNMENTS

RESULT 1
PCT-US95-17111A-121
; Sequence 121, Application PC/TUS9517111A
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and
; TITLE OF INVENTION: Detection of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/17111A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/340,426
; FILING DATE: 14-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 121:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 397 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-17111A-121

Alignment Scores:
Pred. No.: 1,4e-26
Score: 304.00
Matches: 123
Percent Similarity: 44.51%
Best Local Similarity: 35.55%
Query Match: 3.16%
Indels: 80
Gaps: 13

US-09-966-880A-9 (1-5514) x PCT-US95-17111A-121 (1-397)

QY 4441 CTTCCCTCCACAGGACACCACTGATTGCTTTCTTTTCC----- 4394
Db 77 ProSerArgProProLysValLeuAspThrGlyLeuSerThrMetProGlyLeuCysLeu 96
QY 4393 -----ATTTTTGGAGATGGAGTTTCACTGTGTGCCAGGCTGGTCTTGAACCTCT 4340
Db 97 AlaAsnPheCysGlyArgAsnArgValSerLeuMetCysProSerTrpSer-----Pro 114
QY 4339 GAGTTCTAGCAATCTCCCA-COTGGCCTCTCAAGTGTGGGATTACAGGTGTGAGCCT 4281
Db 115 GluLeuLysGlnSerThrCysLeuSerLeuProLysCysTrpAspTrpArgArg--Ala 134
QY 4280 CTATGCTGCCCTGCTTCTTCTGTACTA-----TAAGTACTTCTTCATTCTTCT 4230
Db 134 laValProGlyLeuPheLeuLeuPhePheLeuArgHisArgCysProThrLeuThrGln 154
QY 4229 AGAATTGCTGATG-----CATACATTATCTCTCTCATCTGCTGACTCTTTTCAATTA 4176
Db 154 spGluValGlnTrpCysAspHisSerSerLeuGlnProSerThrLeuArgSerSerIle- 173
QY 4175 GCCTCATGACTTGGGATTCATCCCAAGTAGTCTTCTTGTACTGCAGGGTGGTAT 4116
Db 174 -----LeuLeuProGlnProProLysValAla- 182
QY 4115 TCATTGCTGATAGACATTTTGTGTTTGTTCATCTGCTGATGATTTGGAT 4056
Db 183 --GlyThrLysAspMetHisHisTyThrTrpLeulle-----PheIlePhe---- 197
QY 4055 AGTCTCTACTTGGGGTTTGTG----- 4033
Db 198 -----IlePheAsnPheLeuArgGlnSerLeuAsnSerValThrGlnAlaGlyValG 215
QY 4032 -----AATAAGTACTATGACATTCACGTACCGTCTTT----- 3997
Db 215 lnTrpArgAsnLeuGlySerLeuGlnProLeuProProGlyPheLysLeuPheSerCysP 235
QY 3996 -----GTGCTGATAT 3987
Db 235 roSerLeuLeuSerSerTrpAspTrpArgProProArgProProArgLeuAlaAsnPhePheValP 255
QY 3986 ACATTGAGAAATGATTTTAAATAGTAGTCAGGTCACAGCCCTGACTTTAAAGTT 3927
Db 255 heLeuValGluMetGlyPheHisHisValArgGlnValAspAlaArgSerLeuAspLeuV 275
QY 3926 AAATACCAAAATATCAGCTAACTAAGTAGCTCATCTAGGACACAGCTCTTTGTAGTTG 3867
Db 275 alileCysLeuProArgProProLysValLeuGlyLeu--GlnAspValThrProThrA 294
QY 3866 TTTCCATTATTTTAAATTTTCTTATTTTGTAGGCAAGGCTGCTGGTCTGCTCACC 3807
Db 294 laArgPro-IlePheAsnPheCys-----LeuPheGluMetGluSerHisSerValThr 311
QY 3806 CAGGCTGAGTGCAGTGGCGCAAGCAGCTCCTCAGTGTGAGCCTTGACCTCTGGGGTCCAA 3747
Db 312 GlnAlaGlyValGlnTrpProAsnLeuGlySerLeuGlnProLeuProProGlyLeuLys 331

; SEQUENCE CHARACTERISTICS:
; LENGTH: 397 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-17111A-121

Alignment Scores:
Pred. No.: 1e-24
Score: 286.00
Matches: 132
Percent Similarity: 65.42%
Best Local Similarity: 56.07%
Query Match: 2.97%
Indels: 5
Gaps: 2

US-09-966-880A-9 (1-5514) x US-09-605-785-573 (1-132)

QY 3851 AATTTCCTCTTATTTTGTGAGCAAGGCTGCTGTACCCAGCCTGGAGTGCAG 3792
Db 27 AsnPhe-----PhePheLeuArgGlnGluSerGlyProValAlaGlnAlaGlyValGln 44
QY 3791 TGGCGCAGCACTGCTCAGTCAGCCTGACCTCCTGGGCTCCAAAGATTCTCTGCGCTC 3732
Db 45 TrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeu 64
QY 3731 AGCCC-CCTAGTAGTGGGACACAGGACATGCACCATGCCCGGCTAATTTTGT--- 3676
Db 65 SerLeuProHisSerTrpAspHisArgTyrrAlaProProHisLeuAlaAsnPheCysSer 84

```

```
QY 3675 ----TATAGAAACAGGGTTTGGCCATGTTGCTCAGGCTGGTCTTGAACTCCAGGGCTCAAG 3619
      |||::: ||| |||||||||||||||| |||||||||||:::
Db 85 PheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerGlyThrProGlyLeuGln 104
QY 3618 CGATCCACCACCTCAGCTCCAGAGTGTGGGATTACAGGGCTGAGCCACACACACCTG 3559
      ::||| ||| ||||||::| ||||| ::| |||||
Db 105 GlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTrpArgHisLysProHisPro 124
QY 3558 GCTGTCTGTCTTCACTTAAAT 3538
      ||| |||||
Db 125 AlaCysHisIleLeuLeuAsn 131

RESULT 3
US-09-605-785-573
; Sequence 573, Application US/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605,785
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 573
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-605-785-573

Alignment Scores:
Pred. No.: 4,61e-21 Length: 132
Score: 255.00 Matches: 50
Percent Similarity: 70.13% Conservative: 4
Best Local Similarity: 64.94% Mismatches: 23
Query Match: 2.64% Indels: 0
DB: Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-605-785-573 (1-132)

QY 5283 TTTTTHTTTGGAGTGGAGTTTGGTCTGTGTGCCATCGTCGAGTGAAGGCATGAY 5342
      ||||||| ||| ||| |||||||::| |||||
Db 28 PhePheLeuArgGlnGlnSerGlyProValAlaGlnAlaGlyValGlnTrpHisAsp 47
QY 5343 CATAGCTCACTGCAACCTCCACCTCTCGGTTCACGAAAGCTGTGCGCTCAGCCCTCCG 5402
      ||||||| ||| ||| |||||
Db 48 LeuSerSerLeuGlnProLeuProHisArgPheLeuGlnPheSerCysLeuSerLeuPro 67
QY 5403 GGTAGTGGATTACAGGCCGCCACACACACATCGTGGCTAAATGTTTGTATTTTAGTAGA 5462
      |||||::| ||||| ||| ||| |||
Db 68 HisSerTrpAspHisArgTyrAlaProProHisLeuAlaAsnPheCysSerPheSerArg 87
QY 5463 GATGGGGTTTACCATTGTTGGCCAGGCTGGTCTCAACCTCCTGACCTCAGA 5513
      ||||||| ||| |||||||::| |||||
Db 88 AspGlyValSerLeuCysSerGlyTrpSerLysThrProGlyLeuGln 104
```

```
RESULT 4
US-09-227-357-193
; Sequence 193, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
```


QY	4447	TGCAGTCCTTCCCTCC	CCACAGGACAACCACTGATTTCCTTTTCCCATTTT	4398
Db			:::	
Db	80	AlaSerGlnSerAlaArgTyrArgThrGlyHisAlaArgLeuCysLeuAlaAsn	Phe 99	
QY	4387	GTG---GACAATGGAGTTTCACTGTGTGGCCAGCGCTGGTCTTGAACCTCCTGAGTTCTAG	4331	
Db				
Db	100	CysGlyArgAsnArgValSerLeuMetCysProSerTrpSer	-----ProGluLeuLys 117	
QY	4330	CAATCTTCCCA-CGTGGCCTCTCAAGGCTGGGATTAC	-----4293	
Db				
Db	118	GlnSerThrCysLeuSerLeuProLysCysTrpAspTyrArgArgAlaAlaValProgly	137	
QY	4292	-----AGGTGTGAGCTCTATCCCTGCCTGATTTGCT	4260	
Db				
Db	138	LeuPheLeuLeuPheLeuArgHisArgCysProThrLeuThrGlnAspGluValGln	157	
QY	4259	TTCTCTAACTAATGTTACTT	-----TCA 4236	
Db		:::		
Db	158	TrpCysAspHisSerSerLeuGlnProSerThrProGluLeuLysHisProProAlaSer	177	
QY	4235	TTTTCTAGAATTTCTGATGGCATA	-----CATTTATCTC---CTCCTTCATGTC 4191	
Db				
Db	178	AlaSerGlnValAla--GlyThrLysaspMethHisTyrThrTrpLeuPheLeu	196	
QY	4190	TCACTCTTTTCATTAGCGTCATGACTTTGGGATTCATCCCAAGTACGTTCTTCTCTTGT	4131	
Db		:::	:::	
Db	197	PheIlePheAsnPheLeuArg	-----GlnSerLeuAsnSerValThrGlnAla 212	
QY	4130	ACTGCAAGGCTGGTATTCATTGTCATGATAGCACACT	-----TWGTTT 4086	
Db		:::	:::	
Db	213	GlyValGlnTrpArgAsnLeuGlySerLeuGlnProLeuProGlyPheLysLeuPhe	232	
QY	4085	ATTGCT---TCACCTGTGTGGATAT	-----TTGGATGATCTC 4050	
Db			:::	
Db	233	SerCysProSerLeuLeuSerSerTrpAspTyrArgArgProProArgLeuAlaAsnPhe	252	
QY	4049	TACTTTGGGGTTTGTGAATAAAGGTACTATGAACATTACAGTACCAGCTTTGTGTGGA	3990	
Db		253	-----PheValPheLeuValGluMetGly 260	
QY	3989	TATACATGTG	-----AGAATTGATATTTTAATATGTTAGTCAGGTGCAGACACCTGACT 3936	
Db				
Db	261	PheThrMetPheAlaArgLeu	-----IleLeuIle-SerGlyProCysAspLeuPro-----277	
QY	3935	TTAAAGCTTAAATACCAAAATATACGCCTAACTAACTAGCTCATCTAGCTAGGACACACGCTC	3876	
Db		278	-----AlaSerAlaSerGlnSerAlaGlyIle 286	
QY	3875	TTGTAGTGTGTTTCCATTT	-----ATTTTTAATTTTCTCTTATTTTTTTTGAGCC 3826	
Db				
Db	286	eThrGlyValSerHisHisAlaArgLeuIlePheAsnPheCys	-----LeuPheGluMe 304	
QY	3825	AAGTCTGTGCTGTGCACCCAGCGCTGAGTGGCGCAAGCACTGCTCAGTCAGCCAGC	3766	
Db				
Db	304	tGluSerHisSerValThrGlnAlaGlyValGlnTrpProAsnLeuGlySerLeuGlnPr	324	
QY	3765	TTGACCTCCNGGCTCCAAAGATTCTCCTGCCTCAGCCC	-CCTAGTAGCTGGACCAACAG 3707	
Db				
Db	324	oLeuProGlyLeuLysArgPheSerCysLeuSerLeuProSerSerTrpAspTyrG	344	
QY	3706	GCACATGCCACCATGCCCGCTAAATTTTGT	-----TATAGAAACAGGGTTTGGCCATG 3653	
Db				
Db	344	yHisLeuProHisProAlaAsnPheCysIlePheIleArgGlyValSerProty	364	
QY	3652	TTGCTCAGGCTGGTCTGTGAACCTCCAGGGCTCAAG	3619	
Db				
Db	364	rLeuSerGlyTrpSerGlnThrProAspLeuArg	375	

```

1  GENERAL INFORMATION:
2  APPLICANT: de la Monte, Suzanne
3  APPLICANT: Wands, Jack R.
4  TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
5  TITLE OF INVENTION: of Alzheimer's Disease
6  NUMBER OF SEQUENCES: 121
7  CORRESPONDENCE ADDRESS:
8  ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
9  STREET: 1100 New York Avenue, Suite 600
10 CITY: Washington
11 STATE: D.C.
12 COUNTRY: U.S.A.
13 ZIP: 20005-3934
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: Floppy disk
16 COMPUTER: IBM PC compatible
17 OPERATING SYSTEM: PC-DOS/MS-DOS
18 SOFTWARE: PatentIn Release #1.0, Version #1.25
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/08/450,673C
21 FILING DATE: 30-MAY-1995
22 CLASSIFICATION: 530
23 ATTORNEY/AGENT INFORMATION:
24 NAME: Ludwig, Steven R.
25 REGISTRATION NUMBER: 36,203
26 REFERENCE/DOCKET NUMBER: 0609.3840004
27 TELECOMMUNICATION INFORMATION:
28 TELEPHONE: (202) 371-2600
29 TELEFAX: (202) 371-2540
30 INFORMATION FOR SEQ ID NO: 121:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH: 375 amino acids
33 TYPE: amino acid
34 TOPOLOGY: linear
35 MOLECULE TYPE: protein
36 US-08-450-673C-121
37
38 Alignment Scores:
39 Pred. NO.: 1.12e-18 Length: 375
40 Score: 237.00 Matches: 116
41 Percent Similarity: 41.40% Conservative: 38
42 Best Local Similarity: 31.18% Mismatches: 123
43 Query Match: 2.46% Indels: 98
44 DB: 2 Gaps: 20
45
46 US-09-966-880A-9 (1-5514) x US-08-450-673C-121 (1-375)

```



```
Db 158 TrpCysAspHisSerSerLeuGlnProSerThrProGluIleLeuHisProProAlaSer 177
Qy 4235 TTTTCTAGAAATTCGTATGGCATA-----CATTATCTC--CTCTTCATGTC 4191
Db 178 AlSerGlnValAlaA---GlyThrLysAspMethHisHisThrTrpLeuIlePheHelle 196
Qy 4190 TGACTCTTTCTTAGCGTCATGACTTTTGGGATTATCCCAAGTAGCTTCTCTCTTGT 4131
Db 197 PheIlePheAsnPheLeuArg-----GlnSerLeuAsnSerValThrGlnAla 212
Qy 4130 ACTCAGGGTGGTATTCATTGCATGGATAGACACATT-----TTGTTT 4086
Db 213 GlyValGlnTrpArgAsnLeuGlySerLeuGlnProLeuProGlyPheLysPheLeuPhe 232
Qy 4085 ATTTGT---TCACTTGTTGATGGATAT-----TTGGATAGTCTC 4050
Db 233 SerCysProSerLeuLeuSerSerTrpAspTrpArgProProArgLeuAlaAsnPhe 252
Qy 4049 TACTTTGGGTTTGTGTAATAAGGTACTATGAACATTACGTATCCAGTCTTTGTGGGA 3990
Db 253 -----PheValPheLeuValGluMetGly 260
Qy 3989 TATACATTG-----AGAAATTGATATTTTATATATAGTCAGGTGCACACCTGACT 3936
Db 261 PheThrMetPheAlaArgLeu---IleLeuIle-SerGlyProCysAspLeuPro----- 277
Qy 3935 TTAAGAGTTAAATACCAAAATATCAGCCTAATCAGCTAAGTAGCTCATCTAGGCAACACGCTCT 3876
Db 278 -----AlaSerAlaSerGlnSerAlaGlyIle 286
Qy 3875 TTGTAGTGTTCATT-----ATTTTAATTTTCTTCTTATTTTGTAGGC 3826
Db 286 eThrGlyValSerHisAlaArgLeuIlePheAsnPheCys-----LeuPheGluMe 304
Qy 3825 AAGTCTGGTCTCTCACCAGGTGGAGTGCAGTGGCGAAGCACTCTCAGTGCAGCC 3766
Db 304 tGluSerHisSerValThrGlnAlaGlyValGlnTrpProAsnLeuGlySerLeuGlnPr 324
Qy 3765 TTGACCTCTGGGTCCAAAGATTCTCTGCTCAGCCC-CCTAGTAGCTGGGACCCAG 3707
Db 324 oLeuProGlyLeuLysArgPheSerCysLeuSerLeuProSerTrpAspTrpGly 344
Qy 3706 GCACATGCCACCATGCCCGGCTAAATTTTGT-----TATAGAACAGGGTTTGGCCATG 3653
Db 344 yHisLeuProProHisProAlaAsnPheCysIlePheIleArgGlyGlyValSerProTy 364
Qy 3652 TTGCTCAGGCTGGTCTTGAACCTCAGGCTCAAG 3619
Db 364 rLeuSerGlyTrpSerGlnThrProAspLeuArg 375

RESULT 9
US-08-454-557C-121
; Sequence 121, Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

Db 158 TrpCysAspHisSerSerLeuGlnProSerThrProGluIleLeuHisProProAlaSer 177
Qy 4235 TTTTCTAGAAATTCGTATGGCATA-----CATTATCTC--CTCTTCATGTC 4191
Db 178 AlSerGlnValAlaA---GlyThrLysAspMethHisHisThrTrpLeuIlePheHelle 196
Qy 4190 TGACTCTTTCTTAGCGTCATGACTTTTGGGATTATCCCAAGTAGCTTCTCTCTTGT 4131
Db 197 PheIlePheAsnPheLeuArg-----GlnSerLeuAsnSerValThrGlnAla 212
Qy 4130 ACTCAGGGTGGTATTCATTGCATGGATAGACACATT-----TTGTTT 4086
Db 213 GlyValGlnTrpArgAsnLeuGlySerLeuGlnProLeuProGlyPheLysPheLeuPhe 232
Qy 4085 ATTTGT---TCACTTGTTGATGGATAT-----TTGGATAGTCTC 4050
Db 233 SerCysProSerLeuLeuSerSerTrpAspTrpArgProProArgLeuAlaAsnPhe 252
Qy 4049 TACTTTGGGTTTGTGTAATAAGGTACTATGAACATTACGTATCCAGTCTTTGTGGGA 3990
Db 253 -----PheValPheLeuValGluMetGly 260
Qy 3989 TATACATTG-----AGAAATTGATATTTTATATATAGTCAGGTGCACACCTGACT 3936
Db 261 PheThrMetPheAlaArgLeu---IleLeuIle-SerGlyProCysAspLeuPro----- 277
Qy 3935 TTAAGAGTTAAATACCAAAATATCAGCCTAATCAGCTAAGTAGCTCATCTAGGCAACACGCTCT 3876
Db 278 -----AlaSerAlaSerGlnSerAlaGlyIle 286
Qy 3875 TTGTAGTGTTCATT-----ATTTTAATTTTCTTCTTATTTTGTAGGC 3826
Db 286 eThrGlyValSerHisAlaArgLeuIlePheAsnPheCys-----LeuPheGluMe 304
Qy 3825 AAGTCTGGTCTCTCACCAGGTGGAGTGCAGTGGCGAAGCACTCTCAGTGCAGCC 3766
Db 304 tGluSerHisSerValThrGlnAlaGlyValGlnTrpProAsnLeuGlySerLeuGlnPr 324
Qy 3765 TTGACCTCTGGGTCCAAAGATTCTCTGCTCAGCCC-CCTAGTAGCTGGGACCCAG 3707
Db 324 oLeuProGlyLeuLysArgPheSerCysLeuSerLeuProSerTrpAspTrpGly 344
Qy 3706 GCACATGCCACCATGCCCGGCTAAATTTTGT-----TATAGAACAGGGTTTGGCCATG 3653
Db 344 yHisLeuProProHisProAlaAsnPheCysIlePheIleArgGlyGlyValSerProTy 364
Qy 3652 TTGCTCAGGCTGGTCTTGAACCTCAGGCTCAAG 3619
Db 364 rLeuSerGlyTrpSerGlnThrProAspLeuArg 375

RESULT 10
US-08-340-426D-121
; Sequence 121, Application US/08340426D
; Patent No. 5948634
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/340,426D
; FILING DATE: 14-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
```

```
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 375 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-340-426D-121

Alignment Scores:
Pred. No.: 1.47e-18 Length: 375
Score: 236.00 Matches: 49
Percent Similarity: 65.00% Conservative: 3
Best Local Similarity: 61.25% Mismatches: 27
Query Match: 2.44% Indels: 1
DB: 2 Gaps: 0

US-09-966-880A-9 (1-5514) x US-08-340-426D-121 (1-375)
QY 5276 TTTTGTGAGATGGAGTTTGTGTTGGCCCATGCTGCGAGTGAATG 5335
   ||| ||| ||||||||| ||| ||| |||||||||:|
Db 297 PheAsnPhcCysLeuPheGluMetGluSerHisSer-ValThrGlnAlaGlyValGlnTr 316
   ||| ||| ||||||||| ||| ||| |||||||||
QY 5336 GCATGAYCATAGCTCACTGCAACCTCCACTCTGGTTTCAAGCAAAGCTGTCGCCTCAG 5395
   ||| ||| ||||||||| ||| ||| |||||||||
Db 316 pProAsnLeuGlySerLeuGlnProLeuProGlyLeuLysArgPheSerCysLeuSe 336
   ||| ||| ||||||||| ||| ||| |||||||||
QY 5396 CCTCCCGGTAGATGGATTACAGCGCCGCCACACACCTCGGCTAAATGTTGTTATTTT 5455
   ||| ||| ||||||||| ||| ||| |||||||||
Db 336 rLeuProSerSerTrpAspTrpGlyHisLeuProHisProAlaAsnPhcCysLeuPh 356
   ||| ||| ||||||||| ||| ||| |||||||||
QY 5456 TAGTAGAGATGGGTTTCCACCATGTTGGCCAGGCTGCTCAAACTCTCGACCTCAGA 5513
   ||| ||| ||||||||| ||| ||| |||||||||
Db 356 eileArgGlyGlyValSerProTyLeuSerGlyTrpSerGlnThrProAspLeuArg 375

RESULT 12
PCT-US95-17111A-121
; Sequence 121, Application PC/TUS9517111A
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and
; TITLE OF INVENTION: Detection of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/17111A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/340,426
; FILING DATE: 14-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 397 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 375 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-450-673C-121

Alignment Scores:
Pred. No.: 1.47e-18 Length: 375
Score: 236.00 Matches: 49
Percent Similarity: 65.00% Conservative: 3
Best Local Similarity: 61.25% Mismatches: 27
Query Match: 2.44% Indels: 1
DB: 2 Gaps: 0

US-09-966-880A-9 (1-5514) x US-08-340-426D-121 (1-375)
QY 5276 TTTTGTGAGATGGAGTTTGTGTTGGCCCATGCTGCGAGTGAATG 5335
   ||| ||| ||||||||| ||| ||| |||||||||:|
Db 297 PheAsnPhcCysLeuPheGluMetGluSerHisSer-ValThrGlnAlaGlyValGlnTr 316
   ||| ||| ||||||||| ||| ||| |||||||||
QY 5336 GCATGAYCATAGCTCACTGCAACCTCCACTCTGGTTTCAAGCAAAGCTGTCGCCTCAG 5395
   ||| ||| ||||||||| ||| ||| |||||||||
Db 316 pProAsnLeuGlySerLeuGlnProLeuProGlyLeuLysArgPheSerCysLeuSe 336
   ||| ||| ||||||||| ||| ||| |||||||||
QY 5396 CCTCCCGGTAGATGGATTACAGCGCCGCCACACACCTCGGCTAAATGTTGTTATTTT 5455
   ||| ||| ||||||||| ||| ||| |||||||||
Db 336 rLeuProSerSerTrpAspTrpGlyHisLeuProHisProAlaAsnPhcCysLeuPh 356
   ||| ||| ||||||||| ||| ||| |||||||||
QY 5456 TAGTAGAGATGGGTTTCCACCATGTTGGCCAGGCTGCTCAAACTCTCGACCTCAGA 5513
   ||| ||| ||||||||| ||| ||| |||||||||
Db 356 eileArgGlyGlyValSerProTyLeuSerGlyTrpSerGlnThrProAspLeuArg 375

RESULT 11
US-08-450-673C-121
; Sequence 121, Application US/08450673C
; Patent No. 5948888
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,673C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
```

PCT-US95-17111A-121

Alignment Scores:
Pred. No.: 9, 01e-17 Length: 397
Score: 221.00 Matches: 49
Percent Similarity: 68.75% Conservative: 6
Best Local Similarity: 61.25% Mismatches: 23
Query Match: 2.29% Indels: 2
DB: 5 Gaps: 0

US-09-966-880A-9 (1-5514) x PCT-US95-17111A-121 (1-397)

QY 5276 TTTTCTTTTGTGAGATGGAGTTTGGCTGTGTCGCCATGCTGGAGTGAATG 5335
Db 298 PheAsnPheCysLeuPheGluMetGluSerHisSer-ValThrGlnAlaGlyValGlnTr 317
QY 5336 GCATGAYCATACCTCACTCCACCTCCACCTCCTGGGTTCAGCAAGCAAGTGTGCGCTCAG 5395
Db 317 pProAsnLeuGlySerLeuGlnProLeuProGlyLeuLysArgPheSerCysLeuSe 337
QY 5396 CCTCCGGGTAGATGGATTACAGGCGCCACACACTCGGCTAATGTTGTATT 5455
Db 337 rLeuProSerSerTrpAspTyr-GlyHisLeuHisHisThrProLeuIlePheValPheS 357
QY 5456 TAGTAGAGATGGGGTTTACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCA 5511
Db 357 erLeuGluAlaGlyPheHisHisIleCysGlnAlaGlyLeuLysLeuLeuThrSer 375

RESULT 13

US-09-058-489-22
; Sequence 22, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: WHI97-08pA
; CURRENT APPLICATION NUMBER: US/09/058,489
; EARLIER FILING DATE: 1998-04-10
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22
; LENGTH: 1079
; TYPE: PRT
; ORGANISM: Human
US-09-058-489-22

Alignment Scores:
Pred. No.: 1, 6e-16 Length: 1079
Score: 221.00 Matches: 53
Percent Similarity: 72.62% Conservative: 8
Best Local Similarity: 63.10% Mismatches: 20
Query Match: 2.30% Indels: 4
DB: 3 Gaps: 1

US-09-966-880A-9 (1-5514) x US-09-058-489-22 (1-1079)

QY 3806 CAGGTGGAGTGCAGTGGCGCAAGCAGCTGCTCAGTCAGCAGCTTGACCTCTGGGTCCAA 3747
Db 995 ArgAlaGlyMetGlnTrpCysAspLeuSerLeuGlnProProGlyPheLys 1014
QY 3746 AGATTCTCTGCTCAGCCC-CCTAGTAGCTGGGACCAGCAGCACATGCCACCATGCCCG 3688
Db 1015 ArgPheSerHisLeuSerLeuProAsnSerTrpAsnTyrArgHisLeuProSerCysPro 1034
QY 3687 GCTAATTTTGTATATA-----GAAACAGGGTTTCCCATGTTGCTCAGGCTGGCTTCA 3634
Db 1035 ThrAsnPheCys-IlePheValGluThrGlyPheHisHisValGlyGlnAlaCysLeuGl 1054

QY 3633 ACTCCAGGCTCAAGCGATCCACCCACCTCAGCCTCCCGAGAGTGTGGGATTACAGCGGT 3574
Db 1054 uLeuLeuThrSerGlyGlyLeuLeuAlaSerAlaSerGlnSerAlaGlyIleThrGlyVa 1074
QY 3573 GAGCCACCAC 3564
Db 1074 lSerHisHis 1077

RESULT 14

US-09-058-489-22
; Sequence 22, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: WHI97-08pA
; CURRENT APPLICATION NUMBER: US/09/058,489
; EARLIER FILING DATE: 1998-04-10
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22
; LENGTH: 1079
; TYPE: PRT
; ORGANISM: Human
US-09-058-489-22

Alignment Scores:

Pred. No.: 1, 85e-15 Length: 1079
Score: 212.00 Matches: 44
Percent Similarity: 69.01% Conservative: 5
Best Local Similarity: 61.97% Mismatches: 21
Query Match: 2.19% Indels: 1
DB: 3 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-058-489-22 (1-1079)

QY 5301 AGTTTGGTCTTGTGTCGCCATGCTGGAGTGAATGCATGAYCATAGCTCAGTGCACCT 5360
Db 989 SerPheGlnGluSerLeuArgAlaGlyMetGlnTrpCysAspLeuSerLeuGlnPro 1008
QY 5361 CCACCTCTCGGTTCAGCAAGCTGTCGCTCAGCTCCCGGTAGATGGGATTACAGG 5420
Db 1009 ProProGlyPheLysArgPheSerHisLeuSerLeuProAsnSerTrpAsnTyrArg 1028
QY 5421 CGCCACACACACCTCGGCTAATGTTGTTATTTTAGTAGAGATGGGTTTCCACCATGT 5480
Db 1029 HisLeuProSerCysProThrAsnPheCysIlePhe-ValGluThrGlyPheHisIleVa 1048
QY 5481 TGGCCAGGTGGTCTCAAACTCCTGACCTCA 5511
Db 1048 lGlyGlnAlaCysLeuGluLeuLeuThrSer 1058

RESULT 15

US-09-227-357-577
; Sequence 577, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; EARLIER FILING DATE: 1999-01-08
; EARLIER FILING DATE: 1998-07-07
; EARLIER FILING DATE: 1998-07-07
; EARLIER FILING DATE: 1997-07-08
; EARLIER FILING DATE: 1997-07-08
; EARLIER FILING DATE: 1997-07-08

```

; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 577
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-227-357-577

```

```

Alignment Scores:
Pred. No.: 1.53e-14 Length: 56
Score: 198.00 Matches: 35
Percent Similarity: 71.70% Conservative: 3
Best Local Similarity: 66.04% Mismatches: 15

```

```

Query Match: 2.05% Indels: 0
DB: 4 Gaps: 0
US-09-966-880A-9 (1-5514) x US-09-227-357-577 (1-56)
QY 5355 CAACCTCCACCTCCTGGGTTCAAGCAAGCTCTCGCCTCAGCCTCCCGGGTAGATGGGAT 5414
   :::::::::::::::::::
Db 4 GluProProProProArgPheLysArgPheSerCysLeuSerLeuLeuSerTrpAsp 23
QY 5415 TACAGGCGCCACACACACTCGGCTAATGTTGTTATTTTATAGTAGATGGGTTTCA 5474
   :::::::::::::::::::
Db 24 TyrArgArgAlaProProHisValAlaIlePheCysThrLeuSerArgAspGlyValLeu 43
QY 5475 CCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAGA 5513
   :::::::::::::::::::
Db 44 ProHisTrpProGlyTrpSerGlnThrProAspLeuLys 56

```

Search completed: June 14, 2003, 19:00:00
Job time : 91.1083 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:26:00 ; Search time 177.542 Seconds
(without alignments)
5641.044 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 9659
Sequence: 1 acagacgaatacataggtccca.....tcaaaactcctgacctcagag 5514

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 817286

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPto_spool/US09966880/runat_14062003_175526_10413/app_query.fasta_1.9493
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09966880_0CGN_1_1_105_@runat_14062003_175526_10413
-NCPu=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=1 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_AA.*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PTC_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	345.5	3.6	361 10	US-09-995-494-107
2	309	3.2	93 9	US-09-764-891-5337
3	289.5	3.0	107 9	US-09-989-920-233
4	287	3.0	126 9	US-10-082-830-207

c 5	286	3.0	94	9	US-10-001-835-135	Sequence 135, App
c 6	286	3.0	132	9	US-10-012-896-573	Sequence 573, App
c 7	286	3.0	132	9	US-09-895-793-573	Sequence 573, App
c 8	286	3.0	132	9	US-09-895-814-573	Sequence 573, App
c 9	286	3.0	132	10	US-09-759-143-573	Sequence 573, App
c 10	286	3.0	132	10	US-09-780-669-573	Sequence 573, App
c 11	286	3.0	132	10	US-09-822-827-573	Sequence 573, App
c 12	286	3.0	135	9	US-10-012-896-884	Sequence 884, App
c 13	286	3.0	135	9	US-09-895-793-884	Sequence 884, App
c 14	286	3.0	135	9	US-09-895-814-884	Sequence 884, App
c 15	286	3.0	135	10	US-09-759-143-884	Sequence 884, App
c 16	286	3.0	135	10	US-09-780-669-884	Sequence 884, App
c 17	286	3.0	135	10	US-09-822-827-884	Sequence 884, App
c 18	282.5	2.9	84	9	US-10-001-835-160	Sequence 160, App
c 19	282	2.9	239	10	US-09-800-729-193	Sequence 193, App
c 20	282	2.9	310	10	US-09-800-729-190	Sequence 190, App
c 21	279.5	2.9	100	9	US-10-016-157A-171	Sequence 171, App
c 22	279	2.9	109	9	US-09-989-919-105	Sequence 105, App
c 23	278	2.9	144	9	US-10-011-585A-145	Sequence 145, App
c 24	272	2.8	115	10	US-09-995-494-79	Sequence 79, App1
c 25	270	2.8	183	9	US-09-989-920-245	Sequence 245, App
c 26	267.5	2.8	449	9	US-10-007-280A-140	Sequence 140, App
c 27	265	2.8	118	10	US-09-764-898-157	Sequence 157, App
c 28	264	2.7	107	10	US-10-074-095-510	Sequence 510, App
c 29	264	2.7	107	10	US-09-764-860-510	Sequence 510, App
c 30	263.5	2.7	105	9	US-09-989-919-121	Sequence 121, App
c 31	261	2.7	144	9	US-10-011-585A-145	Sequence 145, App
c 32	258.5	2.7	116	9	US-10-001-835-161	Sequence 161, App
c 33	258.5	2.7	207	9	US-09-866-050A-489	Sequence 489, App
c 34	256.5	2.7	361	10	US-09-995-494-107	Sequence 107, App
c 35	256	2.7	123	9	US-09-764-868-888	Sequence 888, App
c 36	255	2.6	90	9	US-10-001-873-34	Sequence 34, App1
c 37	255	2.6	132	9	US-10-012-896-573	Sequence 573, App
c 38	255	2.6	132	9	US-09-895-793-573	Sequence 573, App
c 39	255	2.6	132	9	US-09-895-814-573	Sequence 573, App
c 40	255	2.6	132	10	US-09-759-143-573	Sequence 573, App
c 41	255	2.6	132	10	US-09-780-669-573	Sequence 573, App
c 42	255	2.6	132	10	US-09-822-827-573	Sequence 573, App
c 43	255	2.6	135	9	US-10-012-896-884	Sequence 884, App
c 44	255	2.6	135	9	US-09-895-793-884	Sequence 884, App
c 45	255	2.6	135	9	US-09-895-814-884	Sequence 884, App

ALIGNMENTS

RESULT 1
US-09-995-494-107
; Sequence 107, Application US/0995494
; Patent No. US20020127578A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Cafferkey, Robert
; APPLICANT: Ali, Shujath
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; APPLICANT: Chen, Sei-Yu
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes
; FILE REFERENCE: DEX-0293
; CURRENT APPLICATION NUMBER: US/09/995,494
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 60/253,176
; PRIOR FILING DATE: 2000-11-27
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 107
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-995-494-107
Alignment Scores:

;
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,802
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 207
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-082-830-207

Alignment Scores:
Pred. No.: 1.6e-19 Length: 126
Score: 287.00 Matches: 70
Percent Similarity: 69.64% Conservative: 8
Best Local Similarity: 62.50% Mismatches: 22
Query Match: 2.98% Indels: 13
DB: Gaps: 3

US-09-966-880A-9 (1-5514) x US-10-082-830-207 (1-126)

QY 3865 TTCCATTATTATTA-----ATTTTCTCTTATTTT 3833
DB 11 PheSerLeuPheLeuValPhePheSerAlaIleSerArgIlePheLeuLeuAla 30
QY 3832 TTGAGCGAAGTCTGTGC---CTGTACCCAGGCTGGAGTGCAGTGGCGAAGCACACTGCT 3776
DB 31 MetSerGlnSerIleMetAlaLeuSerProArgLeuGluCysAsnGlyAlaValSerGly 50
QY 3775 CAGTGCAGCTTGACCTTCCTGGGCTCCAAAGATTCTCTCCCTCAGCCCTCA-GTAGCT 3717
DB 51 HisCysAsnProCysLeuProGlySerSerAspSerProProSerAlaSerGlnValAla 70
QY 3716 GGGACACAGCAGCATCCACCATGCCCGCGCTAATTTTGT-----ATAGAACAGGG 3663
DB 71 GlyIleThrGlyThrCysHisHisAlaArgLeuIlePheValPheLeuValGluMetGly 90
QY 3662 TTTTGCCATTGTGCTCAGGCTGTCTTGAACCTCCAGGCTCAGCCAGCTCCACCCACTCA 3603
DB 91 PheHisValGlyGlnAlaGlyLeuGluLeuLeuThrSerGlyAspLeuProThrSer 110
QY 3602 GCCTCCAGAGTCTGGGATTACAGCGTGGAGCCAC 3567
DB 111 AlaSerGlnSerAlaGlyIleThrGlyValSerHis 122

RESULT 5

US-10-001-835-135
; Sequence 135, Application US/10001835
; Patent No. US20020160387A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Hervé
; APPLICANT: Cafferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chonghua
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro
; FILE REFERENCE: DEX-0277
; CURRENT APPLICATION NUMBER: US/10/001.835
; CURRENT FILING DATE: 2001-11-20
; PRIOR FILING DATE: 2001-11-20
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 135
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-835-135

Alignment Scores:
Pred. No.: 1.8e-19 Length: 94
Score: 286.00 Matches: 63

Percent Similarity: 73.40% Conservative: 6
Best Local Similarity: 67.02% Mismatches: 23
Query Match: 2.97% Indels: 3
DB: Gaps: 1
US-09-966-880A-9 (1-5514) x US-10-001-835-135 (1-94)
QY 3826 CAAGGTCTGTCTGTCTCACCCAGGCTGGAGTGCAGTGGCGAAGCACACTGCTCAGTGCAGC 3767
DB 1 GlnSerLeuThrLeuSerProArgLeuGluCysSerGlyThrValSerAlaHisCysAsn 20
QY 3766 CTTGACCTCTCTGGCTCCAAAGATTCTCTCCCTCAGCCCTCAGTGCAGCC-CTAGTAGCTGGGACCACA 3708
DB 21 LeuHisLeuLeuGlySerSerAspSerProAlaSerValSerAlaValAlaGlyThrThr 40
QY 3707 GGCACATGCCACCATGCCCGCTAATTTTGT-----ATAGAACAGGGTTTGGCCAT 3654
DB 41 GlyValArgHisHisAlaTrpLeuIlePheLeuValGluThrValPheCysHis 60
QY 3653 GTTGCTCAGGCTGTCTTGAACCTCAGGCTCAAAGCGATCCAGCCACCTCAGCCCTCCAGC 3594
DB 61 ValGlyGlnAlaGlyLeuLysLeuLeuThrSerGlyAspProProThrSerAlaSerAla 80
QY 3593 AGTCTGGGATTACAGCGTGGAGCCAGCCACCATCCAGCTGGCTGTCT 3552
DB 81 SerThrGlyIleThrGlyMetSerHisCysAlaTrpProSer 94

RESULT 6

US-10-012-896-573
; Sequence 573, Application US/10012896
; Publication No. US20020183251A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Wantanabe, Yoshihiro
; APPLICANT: Mesgher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012.896
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 573
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-896-573

Alignment Scores:
Pred. No.: 2.05e-19 Length: 132
Score: 286.00 Matches: 60
Percent Similarity: 65.42% Conservative: 10

Best Local Similarity: 56.07% Mismatches: 33
Query Match: 2.97% Indels: 5
DB: 9 Gaps: 2

US-09-966-880A-9 (1-5514) x US-10-012-896-573 (1-132)

```
QY 3851 AATTTTCTCTTATTTTGTAGGCAAGTCGTGCTCACCAGGCTGGAGTGCAG 3792
||||| :||| ||||||||| |||||||||
Db 27 AsnPhe-----PhePheLeuArgGlnGluSerGlyProValAlaGlyValGln 44
QY 3791 TGGCGCAAGCACTGCTCAGTCAGTCAGCTTGCCTCTGGGCTCCAAAGATTCCTCCCTC 3732
||| ||| |||:||||| ||| :|||:|||||
Db 45 TrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeu 64
QY 3731 AGCCC-CCTAGTAGCTGGGACACAGGCACATGCCCATGCCCGCTAATTTTGT--- 3676
||| ||| |||||||||:||||| |||||||
Db 65 SerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeuAlaAsnPheCysSer 84
QY 3675 ---TATAGAACAGGGTTTGGCATGTTGCTCAGGCTGCTTGAACCTCCAGGCTCAAG 3619
|||:| ||| ||||||||| ||| |||||||||
Db 85 PheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLysThrProGlyLeuGln 104
QY 3618 CGATCCACCACCTCAGCTCCAGAGTCTGGGATTACAGCGCTGAGCCACACACCTG 3559
:|||| ||| |||:||||| ||||| :|||:|||||
Db 105 GlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProHisPro 124
QY 3558 GCTGCTGTTTCACTTTAAAT 3538
||| |||||
Db 125 AlaCysHisIleLeuLeuAsn 131
```

RESULT 7

US-09-895-793-573

```
: Sequence 573, Application US/09895793
: Publication No. US20020192763A1
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: APPLICANT: Dillon, Davin C.
: APPLICANT: Mitcham, Jennifer L.
: APPLICANT: Harlocker, Susan L.
: APPLICANT: Jiang, Yuqiu
: APPLICANT: Kalos, Michael D.
: APPLICANT: Retter, Marc W.
: APPLICANT: Stolk, John A.
: APPLICANT: Day, Craig H.
: APPLICANT: Vedvick, Thomas S.
: APPLICANT: Carter, Darrick
: APPLICANT: Li, Samuel X.
: APPLICANT: Wang, Aljun
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Hepler, William T.
: APPLICANT: Henderson, Robert A.
: APPLICANT: Hural, John
: APPLICANT: McNeill, Patricia D.
: APPLICANT: Houghton, Raymond L.
: APPLICANT: Vinals de Bassols, Carlota
: APPLICANT: Foy, Teresa
```

```
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
: FILE REFERENCE: 210121.534C2
: CURRENT APPLICATION NUMBER: US/09/895,793
: CURRENT FILING DATE: 2001-06-29
: NUMBER OF SEQ ID NOS: 982
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 573
: LENGTH: 132
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-895-793-573
```

Alignment Scores:

Pred. No.: 2,05e-19 Length: 132
Score: 286.00 Matches: 60

Percent Similarity: 65.42% Conservative: 10
Best Local Similarity: 56.07% Mismatches: 33
Query Match: 2.97% Indels: 5
DB: 9 Gaps: 2

US-09-966-880A-9 (1-5514) x US-09-895-793-573 (1-132)

```
QY 3851 AATTTTCTCTTATTTTGTAGGCAAGTCGTGCTCACCAGGCTGGAGTGCAG 3792
||||| :||| ||||||||| |||||||||
Db 27 AsnPhe-----PhePheLeuArgGlnGluSerGlyProValAlaGlyValGln 44
QY 3791 TGGCGCAAGCACTGCTCAGTCAGTCAGCTTGCCTCTGGGCTCCAAAGATTCCTCCCTC 3732
||| ||| |||:||||| ||| :|||:|||||
Db 45 TrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeu 64
QY 3731 AGCCC-CCTAGTAGCTGGGACACAGGCACATGCCCATGCCCGCTAATTTTGT--- 3676
||| ||| |||||||||:||||| |||||||
Db 65 SerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeuAlaAsnPheCysSer 84
QY 3675 ---TATAGAACAGGGTTTGGCATGTTGCTCAGGCTGCTTGAACCTCCAGGCTCAAG 3619
|||:| ||| ||||||||| ||| |||||||||
Db 85 PheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLysThrProGlyLeuGln 104
QY 3618 CGATCCACCACCTCAGCTCCAGAGTCTGGGATTACAGCGCTGAGCCACACACCTG 3559
:|||| ||| |||:||||| ||||| :|||:|||||
Db 105 GlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProHisPro 124
QY 3558 GCTGCTGTTTCACTTTAAAT 3538
||| |||||
Db 125 AlaCysHisIleLeuLeuAsn 131
```

RESULT 8

US-09-895-814-573

```
: Sequence 573, Application US/09895814
: Publication No. US20020193296A1
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: APPLICANT: Dillon, Davin C.
: APPLICANT: Mitcham, Jennifer L.
: APPLICANT: Harlocker, Susan L.
: APPLICANT: Jiang, Yuqiu
: APPLICANT: Kalos, Michael D.
: APPLICANT: Retter, Marc W.
: APPLICANT: Stolk, John A.
: APPLICANT: Day, Craig H.
: APPLICANT: Vedvick, Thomas S.
: APPLICANT: Carter, Darrick
: APPLICANT: Li, Samuel X.
: APPLICANT: Wang, Aljun
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Hepler, William T.
: APPLICANT: Henderson, Robert A.
: APPLICANT: Hural, John
: APPLICANT: McNeill, Patricia D.
: APPLICANT: Houghton, Raymond L.
: APPLICANT: Vinals de Bassols, Carlota
: APPLICANT: Foy, Teresa
```

```
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
: FILE REFERENCE: 210121.427C26
: CURRENT APPLICATION NUMBER: US/09/895,814
: CURRENT FILING DATE: 2001-06-29
: NUMBER OF SEQ ID NOS: 990
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 573
: LENGTH: 132
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-895-814-573
```

Alignment Scores:

Pred. No.: 2,05e-19 Length: 132

Score:	286.00	Matches:	60
Percent Similarity:	65.42%	Conservative:	10
Best Local Similarity:	56.07%	Mismatches:	33
Query Match:	2.97%	Indels:	5
DB:	9	Gaps:	2

US-09-966-880A-9 (1-5514) x US-09-895-814-573 (1-132)

3851	Qy	AAATTTCTCTCTATATTTTGTGAGGCAAGGCTCGCTGCTCAACCAGGCTGAGTGCAG	3792
		:	
27	Db	AspPhe-----PhePheLeuArgGlnGlnUserGlyProValAlaGlnAlaGlyValGln	44
		:	
3791	Qy	TGCGCAAGCACTGCCTCAGTGCAGCCTTGACCTCTGGGCTCCAAAGATTCTCTGCCTC	3732
		:	
45	Db	TrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheIysGlnPheSerCysLeu	64
		:	
3731	Qy	AGCCC--CCTAGTAGCTGGGACCAACAGCATGCCACCATGCCGGCTAATTTTCTG---	3676
		:	
65	Db	SerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeuAlaAsnPheCysSer	84
		:	
3675	Qy	---TATAGAAACAGGGTTTGGCATGTTGCTCAGGCTGGTCTTGAATCCAGGGCTCAAG	3619
		:	
85	Db	PheSerArgaspGlyValSerLeuCysCysSerGlyTrpSerIysThrProGlyLeuGln	104
		:	
3618	Qy	CGATCACCACCCTCAGCCTCCACAGAGTGCCTGGATTACAGCGCTGAGCCACCACACCTG	3559
		:	
105	Db	GlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProHisPro	124
		:	
3558	Qy	GCTGCTGTGTTTCACTTTAAAT	3538
125	Db	AlaCysHisIleLeuLeuAsn	131

RESULT 9

```

US-09-759-143-573
; Sequence 573, Application US/09759143
; Patent No. US20020022248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759,143
; CURRENT FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 573
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-143-573

```

DB:	10	Gaps:	2
US-09-966-880A-9	(1-5514)	x	US-09-759-143-573 (1-132)

Qy	3851	AAATTTCTCTCTTATATTTTGTGAGGCAAGCTCTGGCTCTGTCA	CCACCAAGCGCTGAGTGCCAG	37392
		:		
Db	27	AsnPhe-----PhePheLeuArgGlnGluSerGlyProValAlaGln	AlaGlyValGln	44
Qy	3791	TGGCGCAAGCACATGCCTCAGTCGACGCTTGACCTCTCGGGCT	CCAAAAGATTCTCCTGCCTC	37332
		:		
Db	45	TriPHisAspLeuSerSerLeuGlnProLeuProHisArgPheLysGln	PheSerCysLeu	64
Qy	3731	AGCCC-CCTAGTAGCTGGGACCAAGCAGCATGCCACCATGCCGG	CTAATTTTGT---	3676
Db	65	SerLeuProHisSerIrpAspHisArgTyrAlaProProHisLeuAla	AsnPheCysSer	84
Qy	3675	---TATAGAAACAGGGTTTGGCATGTGTCTCAGGCTGGCTTTGA	ACTCCAGGCGCTCAAG	3619
Db	85	PheSerArgAspGlyValSerLeuCysCysSerGlyTyrPheSerGly	ThrProGlyLeuGln	104
Qy	3618	CGATCCACCACTCAGCCCTCCACAGAGTGTGGATTACAGGCTG	AGGCACCAACACCTG	3559
Db	105	GlnSerAlaCysLeuGlyLeuProLysCysIrpGlyTyrArgHisLys	ProHisPro	124
Qy	3558	GCTGCTCTGTTCCTACTTTAAAT	3538	
Db	125	AlaCysHisIleLeuLeuAsn	131	

RESULT 10

```

US-09-780-669-573
; Sequence 573, Application US/09780669
; Patent No. US20020051977A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C24
; CURRENT APPLICATION NUMBER: US/09/780,669
; CURRENT FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 943
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 573
; LENGTH: 132
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-780-669-573

```

```

US-09-966-880A-9 (1-5514) x US-09-780-669-573 (1-132)
QY 3851 AATTTTCTCTTATTTTGTGAGCAAGGCTGTGCTGTCAACCCAGGCTGGAGTGCAG 3792
Db 27 AsnPhe-----PhePheLeuArgGlnGluSerGlyProValAlaGlnAlaGlyValGln 44
QY 3791 TGGCGCAAGCACTGCTCAGTGCAGCTTGACCTCTGGCTCCAAAGATTTCTCTGCCTC 3732
Db 45 TrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeu 64
QY 3731 AGCCC-CCTAGTACCTGGGACACGACATGCCACCATGCCCGGCTAATTTTGT--- 3676
Db 65 SerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeuAlaAsnPheCysSer 84
QY 3675 ---TATGAACACAGGTTTGGCCATGTGCTAGGCTGTGTTGAACCTCCAGGGCTCAAG 3619
Db 85 PheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLysThrProGlyLeuGln 104
QY 3618 CGATCCACCCACTCAGCTCCAGAGTGTGGATACAGGGTGTGAGCCACACACCTG 3559
Db 105 GlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProHisPro 124
QY 3558 GCTGCTGCTTCACTTTAAAT 3538
Db 125 AlaCysHisLeuLeuLeuAsn 131
RESULT 11
US-09-822-827-573
; Sequence 573, Application US/09822827
; Patent No. US20020081680A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.534C1
; CURRENT APPLICATION NUMBER: US/09/822,827
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 573
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-822-827-573
Alignment Scores:
Pred. No.: 2,05e-19 Length: 132
Score: 286.00 Matches: 60
Percent Similarity: 65.42% Conservative: 10
Best Local Similarity: 56.07% Mismatches: 33
Query Match: 2.97% Indels: 5
DB: 10 Gaps: 2
US-09-966-880A-9 (1-5514) x US-09-822-827-573 (1-132)
QY 3851 AATTTTCTCTTATTTTGTGAGCAAGGCTGTGCTGTCAACCCAGGCTGGAGTGCAG 3792
Db 27 AsnPhe-----PhePheLeuArgGlnGluSerGlyProValAlaGlnAlaGlyValGln 44
QY 3791 TGGCGCAAGCACTGCTCAGTGCAGCTTGACCTCTGGCTCCAAAGATTTCTCTGCCTC 3732
Db 45 TrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeu 64
QY 3731 AGCCC-CCTAGTACCTGGGACACGACATGCCACCATGCCCGGCTAATTTTGT--- 3676
Db 65 SerLeuProHisSerTrpAspHisArgTyrAlaProProHisLeuAlaAsnPheCysSer 84
QY 3675 ---TATGAACACAGGTTTGGCCATGTGCTAGGCTGTGTTGAACCTCCAGGGCTCAAG 3619
Db 85 PheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLysThrProGlyLeuGln 104
QY 3618 CGATCCACCCACTCAGCTCCAGAGTGTGGATACAGGGTGTGAGCCACACACCTG 3559

```

Db 85 PheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLysThrProGlyLeuGln 104
QY 3618 CGATCCACCCACCTCCAGAGTGTGGAGTACAGCGTGCAGCCACACCTG 3559
Db 105 GlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProHisPro 124
QY 3558 GCTGCTGTGTTCACTTTAAAT 3538
Db 125 AlaCysHisIleLeuLeuAsn 131

RESULT 13

US-09-895-793-884
; Sequence 884, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895,793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 884
; LENGTH: 135
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-793-884

Alignment Scores:
Pred. No.: 2,07e-19 Length: 135
Score: 286.00 Matches: 60
Percent Similarity: 65.42% Conservative: 10
Best Local Similarity: 56.07% Mismatches: 33
Query Match: 2.97% Indels: 5
DB: Gaps: 2

US-09-966-880A-9 (1-5514) x US-09-895-793-884 (1-135)

QY 3851 AATTTTCTCTTATTTTGTGAGCAAGGTCTGGTCTGTACCCAGGCTGGAGTGCAG 3792
Db 27 AsnPhe-----PhePheLeuArgGlnGluSerGlyProValAlaGlnAlaGlyValGln 44
QY 3791 TGGCGCAAGCACTGCTCAGTCAGCCCTTGACCTCTGGGCTCCAAAGATTCTCTGCCTC 3732
Db 45 TrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeu 64
QY 3731 AGCCC-CCTAGTAGCTGGGACCACAGGCACATGCACCATGCCCGCGCTAATTTTGT--- 3676
Db 65 SerLeuProHisSerTrpAspHisArgTyrAlaProHisLeuAlaAsnPheCysSer 84
QY 3675 ---TATAGAAACAGGGTTTGGCCATGTGTGCTCAGGCTGGTGTGAAGTCCAGGCTCAAG 3619

Db 85 PheSerArgAspGlyValSerLeuCysCysSerGlyTrpSerLysThrProGlyLeuGln 104
QY 3618 CGATCCACCCACCTCCAGAGTGTGGAGTACAGCGTGCAGCCACACCTG 3559
Db 105 GlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProHisPro 124
QY 3558 GCTGCTGTGTTCACTTTAAAT 3538
Db 125 AlaCysHisIleLeuLeuAsn 131

RESULT 14

US-09-895-814-884
; Sequence 884, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 884
; LENGTH: 135
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-814-884

Alignment Scores:
Pred. No.: 2,07e-19 Length: 135
Score: 286.00 Matches: 60
Percent Similarity: 65.42% Conservative: 10
Best Local Similarity: 56.07% Mismatches: 33
Query Match: 2.97% Indels: 5
DB: Gaps: 2

US-09-966-880A-9 (1-5514) x US-09-895-814-884 (1-135)

QY 3851 AATTTTCTCTTATTTTGTGAGCAAGGTCTGGTCTGTACCCAGGCTGGAGTGCAG 3792
Db 27 AsnPhe-----PhePheLeuArgGlnGluSerGlyProValAlaGlnAlaGlyValGln 44
QY 3791 TGGCGCAAGCACTGCTCAGTCAGCCCTTGACCTCTGGGCTCCAAAGATTCTCTGCCTC 3732
Db 45 TrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeu 64
QY 3731 AGCCC-CCTAGTAGCTGGGACCACAGGCACATGCACCATGCCCGCGCTAATTTTGT--- 3676
Db 65 SerLeuProHisSerTrpAspHisArgTyrAlaProHisLeuAlaAsnPheCysSer 84

QY 3675 ---TATAGAAACAGGGTTTCCCATGTTGCTCAGGCTGGTCTTGAACCTCAGGCGCTCAAG 3619
Db 85 PheSerArgAspGlyValSerLeuCysSerGlyTrpSerLysThrProGlyLeuGln 104
QY 3618 CGATCCACCCACCTCCAGCTCCAGAGTGTGGGATTACAGCGCTGAGCCACACACCTG 3559
Db 105 GlnSerAlaCysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProProHisPro 124
QY 3558 GCTGCTGCTGTTTCACTTTAAAT 3538
Db 125 AlaCysHisIleLeuLeuAsn 131

RESULT 15

US-09-759-143-884

; Sequence 884, Application US/09759143

; Patent No. US20020022248A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Dillon, Davin C.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan L.

; APPLICANT: Jiang, Yuqi

; APPLICANT: Henderson, Robert A.

; APPLICANT: Kalos, Michael D.

; APPLICANT: Fanger, Gary R.

; APPLICANT: Retter, Marc W.

; APPLICANT: Stolk, John A.

; APPLICANT: Day, Craig H.

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Carter, Darrick

; APPLICANT: Li, Samuel

; APPLICANT: Wang, Aijun

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Hepler, William

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.427C23

; CURRENT APPLICATION NUMBER: US/09/759,143

; CURRENT FILING DATE: 2001-01-12

; NUMBER OF SEQ ID NOS: 934

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 884

; LENGTH: 135

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-759-143-884

Alignment Scores:

Pred. No.:	2,07e-19	Length:	135
Score:	286.00	Matches:	60
Percent Similarity:	65.42%	Conservative:	10
Best Local Similarity:	56.07%	Mismatches:	33
Query Match:	2.97%	Indels:	5
DB:	10	Gaps:	2

US-09-966-880A-9 (1-5514) x US-09-759-143-884 (1-135)

QY 3851 AATTTTCTCTCTATTTTGGCAAGTCTGCTGCATCCACAGGCTGGAGTGCAG 3792
Db 27 AsnPhe-----PhePheLeuArgGlnGlnSerGlyProValAlaGlnAlaGlyValGln 44
QY 3791 TGGCGCAGCAGCTGCTCAGTCAGCCTTGCCTCCTGGGCTCCAAAGATTCTCCTGCCTC 3732
Db 45 TrpHisAspLeuSerSerLeuGlnProLeuProHisArgPheLysGlnPheSerCysLeu 64
QY 3731 AGCCC-CCTAGTAGCGGACGACGACATGCCACCATGCCCGGCTAATTTTGT--- 3676
Db 65 SerLeuProHisSerTrpAspHisArgTyrAlaProHisLeuAlaAsnPheCysSer 84
QY 3675 ---TATAGAAACAGGGTTTCCCATGTTGCTCAGGCTGGTCTTGAACCTCAGGCGCTCAAG 3619
Db 85 PheSerArgAspGlyValSerLeuCysSerGlyTrpSerLysThrProGlyLeuGln 104

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 11:14:11 ; Search time 13758 Seconds
(without alignments)
11663.968 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 5514
Sequence: 1 acagacgaatacatagtggtcca.....tcaaacctctgacctcagag 5514

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vl.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4771	86.5	11204	9	AB040430 Homo sapi
2	4649	84.3	71132	9	AC092184 Homo sapi
3	84	1.5	1837	9	BC006296 Homo sapi
4	84	1.5	2791	9	AB040431 Homo sapi
c 5	75	1.4	84778	9	AL353650 Human DNA
c 6	75	1.4	123253	9	AC004971 Homo sapi
c 7	75	1.4	168524	2	AL356418 Homo sapi
8	74	1.3	29359	2	AP001083 Homo sapi
9	74	1.3	168079	2	AC024709 Homo sapi
10	74	1.3	172376	2	AL353749 Homo sapi
11	74	1.3	193280	9	AL139113 Human DNA
c 12	74	1.3	215719	9	AC048334 Homo sapi
c 13	73	1.3	173943	2	AC022717 Homo sapi
c 14	73	1.3	195661	2	AC027595 Homo sapi
15	73	1.3	214696	2	AF235101 Homo sapi
c 16	71	1.3	152129	2	AC027416 Homo sapi
c 17	71	1.3	206854	9	AC009974 Homo sapi
c 18	69	1.3	3187	9	AK092318 Homo sapi
19	69	1.3	79419	9	HSJ324017 Human DNA
20	69	1.3	81364	2	AC021072_3 Continuation (4 of
21	69	1.3	109138	9	AC104084 Homo sapi
c 22	69	1.3	127862	9	AL445430 Human DNA
c 23	69	1.3	146574	2	AC023459 Homo sapi
24	69	1.3	154162	2	AC069270 Homo sapi
c 25	69	1.3	157807	9	AC073573 Homo sapi
c 26	69	1.3	164168	9	AL135927 Human DNA
c 27	69	1.3	164179	9	AC007227 Homo sapi
c 28	69	1.3	193717	2	AC068888 Homo sapi
c 29	68	1.2	4285	9	AC093077 Homo sapi
c 30	68	1.2	8429	9	AF350881 Homo sapi
31	68	1.2	39212	9	HSN21F1 Human DNA s
c 32	68	1.2	44890	9	AC005778 Homo sapi
c 33	68	1.2	51012	9	AC025536 Homo sapi
c 34	68	1.2	57779	2	AC115982 Homo sapi
35	68	1.2	90543	9	AC090287 Homo sapi
36	68	1.2	111035	9	AC022405 Homo sapi
c 37	68	1.2	131928	9	AC091738 Homo sapi
c 38	68	1.2	135693	9	HSJ1167H4 Human DNA
c 39	68	1.2	141655	9	HS1093G12 Human DNA
c 40	68	1.2	146568	2	AC022651 Homo sapi
c 41	68	1.2	148103	9	AL445232 Human DNA
c 42	68	1.2	149726	9	AC016394 Homo sapi
c 43	68	1.2	150192	9	AC008686 Homo sapi
c 44	68	1.2	156949	9	AL161652 Human DNA
c 45	68	1.2	157860	9	AL391241 Human DNA

ALIGNMENTS

RESULT 1
AB040430 Homo sapiens AID gene for activation-induced cytidine deaminase,
LOCUS complete cds.
DEFINITION
AB040430.1 GI:9988407
VERSION
AID: activation-induced cytidine deaminase.
KEYWORDS
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
AUTHORS Muto,T., Muramatsu,M., Tanikawa,K. and Honjo,T.
TITLE Isolation, tissue distribution, and chromosomal localization of the

JOURNAL MEDLINE REFERENCE	human activation-induced cytidine deaminase (AID) gene Genomics 68 (1), 85-88 (2000)	Db	421	AATGCACCTGTCAGACTGAGACAGAGAACCAATCAATTAATTGAAGTGTGAGATTTTCTTGGCCT	480
AUTHORS	2 (sites) Revu,P., Muto,T., Levy,Y., Geissmann,F., Plebani,A., Sanal,O., Catalan,N., Forveille,M., Dufourcq-Lagelouse,R., Gennery,A., Tezcan,I., Eroy,F., Kayserili,H., Ugazio,A.G., Brousse,N., Muramatsu,M., Taniguchi,K., Kinoshita,K., Honjo,T., Fischer,A. and Durandy,A.	QY	1071	GAGACTTGGAGGAGGCAAGACACACTCTGGACACCACTATGACAGAGTAAAGAGCGAG	1130
TITLE	Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2)	Db	481	GAGACTTGGAGGAGGCAAGACACACTCTGGACACCACTATGACAGAGTAAAGAGCGAG	540
JOURNAL MEDLINE REFERENCE	Cell 102 (5), 565-575 (2000)	QY	1131	TCTTCGTCGGTGGTGAATGACACTGGCCTTCTCTCAGAGCAAACTCTGAGTAATGAGACTGG	1190
AUTHORS	3 (bases 1 to 11204) Muto,T., Muramatsu,M., Taniwaki,M., Kinoshita,K. and Honjo,T.	Db	541	TCTTCGTCGGTGGTGAATGACACTGGCCTTCTCTCAGAGCAAACTCTGAGTAATGAGACTGG	600
TITLE	Direct Submission Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department of Medical Chemistry, Faculty of Medicine, Yoshida, Sakyo-ku, Kyoto, Kyoto 606-8501, Japan (E-mail:honjo@four.med.kyoto-u.ac.jp, Tel:81-75-753-4371(ex.4371), Fax:81-75-753-4388)	QY	1191	TAGCTATCCCTTTCTCTCATGTAACCTGCTGACTGATAGATGAGCTTGTATCAATATGCA	1250
JOURNAL MEDLINE REFERENCE	Location/Qualifiers 1. .11204	Db	601	TAGCTATCCCTTTCTCTCATGTAACCTGCTGACTGATAGATGAGCTTGTATCAATATGCA	660
AUTHORS	/organism="Homo sapiens" /db_xref="taxon:9606"	QY	1251	TATATATTTTGTGATCTGCTCTCTCTTCTTATTCAGATCTTATACGGTGTGAGCCCAA	1310
TITLE	Join(521..528,6280..6427,7807..8077,8371..8486,8956..9009) /gene="AID"	Db	661	TATATATTTTGTGATCTGCTCTCTCTTCTTATTCAGATCTTATACGGTGTGAGCCCAA	720
JOURNAL	Join(521..528,6280..6427,7807..8077,8371..8486,8956..9009) /gene="AID"	QY	1311	TTCTTTCTGTTTCAGACTTCTCTTTGATTTCCCTCTTTTTCATGTGGCAAAAGAGTAGTG	1370
FEATURES	/codon_start=1 /product="activation-induced cytidine deaminase"	Db	721	TTCTTTCTGTTTCAGACTTCTCTTTGATTTCCCTCTTTTTCATGTGGCAAAAGAGTAGTG	780
source	/protein_id="BAB1270.1"	QY	1371	CGTACAACTGATGATTCGCTCTGAGATTTGTACCATGTTGAACTAATTTATGTAAT	1430
gene	/db_xref="GI:9988408"	Db	781	CGTACAACTGATGATTCGCTCTGAGATTTGTACCATGTTGAACTAATTTATGTAAT	840
CDS	AWGELHNSVRLSRLRLILLPLXYEVDLDRFAFTLGL"	QY	1431	AAATATTAACATAGCAAAATCTTTAGAGACTCAAAATCATGAAAGGTAAATAGCAGTACTGTA	1490
BASE COUNT	3305 a 2273 c 2373 g 3253 t	Db	841	AAATATTAACATAGCAAAATCTTTAGAGACTCAAAATCATGAAAGGTAAATAGCAGTACTGTA	900
ORIGIN		QY	1491	CTAAAGCGGTAGTGTCTAATTTTCGTAATAATTTTCTTAATAATTTCAACAGTAAACAACAT	1550
Query Match	86.5%; Score 4771; DB 9; Length 11204;	Db	901	CTAAAGCGGTAGTGTCTAATTTTCGTAATAATTTTCTTAATAATTTCAACAGTAAACAACAT	960
Best Local Similarity	99.9%; Pred. No. 0;	QY	1551	TGAAGACACACTTTCCTAGGAGGCGTACTGAAATAATTTAGCTATAGTAAAGAAATTT	1610
Matches 4921; Conservative	0; Mismatches 3; Indels 0; Gaps 0;	Db	961	TGAAGACACACTTTCCTAGGAGGCGTACTGAAATAATTTAGCTATAGTAAAGAAATTT	1020
QY	591 AGTTTCAGAGACACTGGGAATATGGGGAATTTAGAGGCTATCTGAGGCTCTTCAACAC	QY	1611	GTAAATTTAGAAATGCCAAGCATTTCTAAATTAATTTGCTTTGAAAGTCACATATGATTTGTGTC	1670
Db	1 AGTTTCAGAGACACTGGGAATATGGGGAATTTAGAGGCTATCTGAGGCTCTTCAACAC	Db	1021	GTAAATTTAGAAATGCCAAGCATTTCTAAATTAATTTGCTTTGAAAGTCACATATGATTTGTGTC	1080
QY	651 AATAACCCAGAGACTATTTAAATGCTCTTTAAAGGTATTTACATAAAATATTACTATTCTC	QY	1671	CATTATAAGGAGACAAATTCATTCAAGCAAGTTATTTAAATGTTAAAGGCCCAATTTGTTAG	1730
Db	61 AATAACCCAGAGACTATTTAAATGCTCTTTAAAGGTATTTACATAAAATATTACTATTCTC	Db	1081	CATTATAAGGAGACAAATTCATTCAAGCAAGTTATTTAAATGTTAAAGGCCCAATTTGTTAG	1140
QY	711 ATGTGCTTTTATTTTGTATCATGATTAATTAAGTATTTAGAGTGTCTACTGCTTACTGCCCTC	QY	1731	GCAGTTAATGGCACATTTTACTTAACTAATCTTTCCCATTTGTTTCAGAGCTAGCTTAACT	1790
Db	121 ATGTGCTTTTATTTTGTATCATGATTAATTAAGTATTTAGAGTGTCTACTGCTTACTGCCCTC	Db	1141	GCAGTTAATGGCACATTTTACTTAACTAATCTTTCCCATTTGTTTCAGAGCTAGCTTAACT	1200
QY	771 TGATCTTTGTAGCTATGAGCATGAGCTGGCTTTTAGAGCAGCAGCCCCAAGGAACC	QY	1791	TACCTCTTAGGTGTGAATTTGGTTAAGGTCTCTCATAATGCTTTTATGTCAGTGTGTTGAT	1850
Db	181 TGATCTTTGTAGCTATGAGCATGAGCTGGCTTTTAGAGCAGCAGCCCCAAGGAACC	Db	1201	TACCTCTTAGGTGTGAATTTGGTTAAGGTCTCTCATAATGCTTTTATGTCAGTGTGTTGAT	1260
QY	831 TAAACATTAAAGCAGAGCTGCCCTCAATGGTTTAAACCTGTGTGACTCTGCCCTATGACAGC	QY	1851	AGGTTATTGTCATAGAACTTATCTTATCTACATTTTATGATTTACTATGGATGTATGAGA	1910
Db	241 TAAACATTAAAGCAGAGCTGCCCTCAATGGTTTAAACCTGTGTGACTCTGCCCTATGACAGC	Db	1261	AGGTTATTGTCATAGAACTTATCTTATCTACATTTTATGATTTACTATGGATGTATGAGA	1320
QY	891 CCCACCCACCATCTTCTACTGGATCCAAATCAGGAGCAAGGCCGTTGGGGTACCTGGTGG	QY	1911	ATAACACCTTAATCTCTTATACTTAACTCAATTTTAACTCCCTTTTATAAAGAACTTACATTAAC	1970
Db	301 CCCACCCACCATCTTCTACTGGATCCAAATCAGGAGCAAGGCCGTTGGGGTACCTGGTGG	Db	1321	ATAACACCTTAATCTCTTATACTTAACTCAATTTTAACTCCCTTTTATAAAGAACTTACATTAAC	1380
QY	951 GGGTGATGCTGTGAGGAGGAGGCCCAAAGGGCAAGCTCAAAATTTGAATGTGAGGGGCC	QY	1971	AGAATAAAGATTTTTTAAATAATATATTTTTTTTGTAGAGACAGGGCTTTAGCCCCAGCCGAG	2030
Db	361 GGGTGATGCTGTGAGGAGGAGGCCCAAAGGGCAAGCTCAAAATTTGAATGTGAGGGGCC	Db	1381	AGAATAAAGATTTTTTAAATAATATATTTTTTTTGTAGAGACAGGGCTTTAGCCCCAGCCGAG	1440
QY	1011 AATGCACCTGTGAGACTGAGACAGAGAACCATCAATTAATTGAAGTGTGAGATTTTCTTGGCCT	QY	2031	GCTGCTCTAGTCTTGGCCCAAGGATCTCTCTGCTGGGCTCCCTAAAGCTGCTGGAA	2090
Db		Db	1441	GCTGCTCTAGTCTTGGCCCAAGGATCTCTCTGCTGGGCTCCCTAAAGCTGCTGGAA	1500
QY		QY	2091	TTATAGACATGAGCCATCACATCCCAATATACAGATAAAGATTTTTTAAATGGAGGATTTAA	2150
		Db	1501	TTATAGACATGAGCCATCACATCCCAATATACAGATAAAGATTTTTTAAATGGAGGATTTAA	1560

QY 2151 TGTCTTTCAGAAAAATTTCTTGAGGTCAGACAAATGTCAAAATGTCCTCCTCAGTTTACACTG 2210
DB TGTCTTTCAGAAAAATTTCTTGAGGTCAGACAAATGTCAAAATGTCCTCCTCAGTTTACACTG 1620
QY 2211 AGATTTTGAACAAGTCTCAGCTATAGGTCCTGTGTGAAGGGTCCATTTGGAATACATTGT 2270
DB AGATTTTGAACAAGTCTCAGCTATAGGTCCTGTGTGAAGGGTCCATTTGGAATACATTGT 1680
QY 2271 TCAAGTAAATGGAAGCAAGGTAAATCAGCAGTTGAAATTCAGAGAAAGACAGAAA 2330
DB TCAAGTAAATGGAAGCAAGGTAAATCAGCAGTTGAAATTCAGAGAAAGACAGAAA 1740
QY 2331 AGGAGAAAGATGAAATTCACAGGACAGAGGAAATATATATATCATTAAGGAGGACAG 2390
DB AGGAGAAAGATGAAATTCACAGGACAGAGGAAATATATATATCATTAAGGAGGACAG 1800
QY 2391 TATCTGTAGAGCTCAATAGTATGGCAAAATGACTTGGTCAGGATTAATTTTAAACCGCT 2450
DB TATCTGTAGAGCTCAATAGTATGGCAAAATGACTTGGTCAGGATTAATTTTAAACCGCT 1860
QY 2451 TGTCTTCTGGTTTCACGGCTGGGATGCAGCTAGGGTTCTGCTCAGGGACACAGCTGT 2510
DB TGTCTTCTGGTTTCACGGCTGGGATGCAGCTAGGGTTCTGCTCAGGGACACAGCTGT 1920
QY 2511 CCAGAGCAGCTGTGAGCTGCAAGCCTGAAACACTCCTCGGTAAAGTCTCTTCTACTCA 2570
DB CCAGAGCAGCTGTGAGCTGCAAGCCTGAAACACTCCTCGGTAAAGTCTCTTCTACTCA 1980
QY 2571 GGACAGAAATGACGAGAACAGGAGCTGGAACAGGCCCTTAACCCAGAGAGGGAAGTAA 2630
DB GGACAGAAATGACGAGAACAGGAGCTGGAACAGGCCCTTAACCCAGAGAGGGAAGTAA 2040
QY 2631 TGGATCAACAAGTTAACTAGCAGGTGAGGTCAGGATCAGCAATTCATTTCACTCTGACTGGTA 2690
DB TGGATCAACAAGTTAACTAGCAGGTGAGGTCAGGATCAGCAATTCATTTCACTCTGACTGGTA 2100
QY 2691 ACATGTGACAGAAACAGTGTAGGCTTATTTGATTTTTCATGTAGAGTAGGACCCAAAATC 2750
DB ACATGTGACAGAAACAGTGTAGGCTTATTTGATTTTTCATGTAGAGTAGGACCCAAAATC 2160
QY 2751 CACCCAAAGTCTTTATCTATGCCATCTCTTATCTATCTATCTATCTATCTATCTATCTATCT 2810
DB CACCCAAAGTCTTTATCTATGCCATCTCTTATCTATCTATCTATCTATCTATCTATCTATCT 2220
QY 2811 TTCTTATGATAAGGCTCTCTCTCCACACACACACACACACACACACACACACACACACACA 2870
DB TTCTTATGATAAGGCTCTCTCTCTCCACACACACACACACACACACACACACACACACACACA 2280
QY 2871 CA 2930
DB CA 2340
QY 2931 CTCTGCCCTTCTCATCTACACAGCCCGCCCAAGGTGATGTATAAAGAGGATTTATGG 2990
DB CTCTGCCCTTCTCATCTACACAGCCCGCCCAAGGTGATGTATAAAGAGGATTTATGG 2400
QY 2991 TAAGAGATGATGCTTATCTGTTTAAACACCTGGGCTCAAGAGAGAAATTTCTTCTCTCT 3050
DB TAAGAGATGATGCTTATCTGTTTAAACACCTGGGCTCAAGAGAGAAATTTCTTCTCTCTCT 2460
QY 3051 GTACTTTAAGCACCTTATGTTGAGCTTATATATACAAAGGGTATATATATGCTA 3110
DB GTACTTTAAGCACCTTATGTTGAGCTTATATATACAAAGGGTATATATATGCTA 2520
QY 3111 ATATAGTAATAGTAATGTTGGTGTACTATGTTAAATACCATAAAAAATTAATATCCTTT 3170
DB ATATAGTAATAGTAATGTTGGTGTACTATGTTAAATACCATAAAAAATTAATATCCTTT 2580
QY 3171 TAAATTAAGCTTAATATTTGATCTTTTATGATTTATCATTTTATGTTTATGTTT 3230
DB TAAATTAAGCTTAATATTTGATCTTTTATGATTTATCATTTTATGTTTATGTTT 2640

QY 3231 TTGATTTTAAAGACAAATCTCACCTGTTTACCAGGCTGAGTGCAGTGGTGAATCA 3290
DB TTGATTTTAAAGACAAATCTCACCTGTTTACCAGGCTGAGTGCAGTGGTGAATCA 2700
QY 3291 TAGCTTTCTCAGTCTTGAACCTCTGGGCTCAAGCAATCTCTCTGCTTGGCTTCCCAAA 3350
DB TAGCTTTCTCAGTCTTGAACCTCTGGGCTCAAGCAATCTCTCTGCTTGGCTTCCCAAA 2760
QY 3351 GTGTGGGATACAGTCAATGAGCCACTGCATCTGSCCTAGGATCCATTTAGATTTAAATAT 3410
DB GTGTGGGATACAGTCAATGAGCCACTGCATCTGSCCTAGGATCCATTTAGATTTAAATAT 2820
QY 3411 GCATTTTAAATTTTAAATATATATGCTTAAATTTTACCTTATGTAATGTTGATCTGCTA 3470
DB GCATTTTAAATTTTAAATATATATGCTTAAATTTTACCTTATGTAATGTTGATCTGCTA 2880
QY 3471 ATAAATCTAGTTGCTGCTGCTTAAAGTTTAAAGTGTCTTCCCAATAGCTTCTAGTGTAG 3530
DB ATAAATCTAGTTGCTGCTGCTTAAAGTTTAAAGTGTCTTCCCAATAGCTTCTAGTGTAG 2940
QY 3531 GGGAGACATTTAAAGTGAACACAGACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3590
DB GGGAGACATTTAAAGTGAACACAGACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3000
QY 3591 ACTCTGGAGGCTGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3650
DB ACTCTGGAGGCTGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3060
QY 3651 AACATGCGAAACCTGTTCTATTAACAAAATTTAGCGGGCATGGTGGCATGGTGGTGGTGGT 3710
DB AACATGCGAAACCTGTTCTATTAACAAAATTTAGCGGGCATGGTGGCATGGTGGTGGTGGT 3120
QY 3711 GGTCTCCAGCTACTAGGGGCTGAGGAGGAGAAATCTTTGGAGCCACAGGAGCTCAAGGCTG 3770
DB GGTCTCCAGCTACTAGGGGCTGAGGAGGAGAAATCTTTGGAGCCACAGGAGCTCAAGGCTG 3180
QY 3771 CACTGAGCAGTGTCTGGCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3830
DB CACTGAGCAGTGTCTGGCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3240
QY 3831 AAAAAATAGAGAAAAATTTAAAAATTAATGGAACAACTACAAAGAGCTGTTGTCCTA 3890
DB AAAAAATAGAGAAAAATTTAAAAATTAATGGAACAACTACAAAGAGCTGTTGTCCTA 3300
QY 3891 GATCAGCTACTTACTAGGCTGATTTTGGTATTTTAACTTTTAAAGTCAGGGTCTGCTCA 3950
DB GATCAGCTACTTACTAGGCTGATTTTGGTATTTTAACTTTTAAAGTCAGGGTCTGCTCA 3360
QY 3951 CCTGCACTACATTTAAAAATATCAATTTCTCAATGTATATCCACACAAAGACTGGTAGCT 4010
DB CCTGCACTACATTTAAAAATATCAATTTCTCAATGTATATCCACACAAAGACTGGTAGCT 3420
QY 4011 GAATGTTCACTAGTACCTTTTATTCACAAACCCCAAGTAGAGACTATCCAAATATCCATC 4070
DB GAATGTTCACTAGTACCTTTTATTCACAAACCCCAAGTAGAGACTATCCAAATATCCATC 3480
QY 4071 AACAAAGTGAACAAATAACAAAAATGCTATATCCATGCAATGAAATACCACCTGCACT 4130
DB AACAAAGTGAACAAATAACAAAAATGCTATATCCATGCAATGAAATACCACCTGCACT 3540
QY 4131 ACRAAGGAAGAGTACTTGGGGATGAATCCCAAGTCATGACGCTAAATGAAAGAGTCA 4190
DB ACRAAGGAAGAGTACTTGGGGATGAATCCCAAGTCATGACGCTAAATGAAAGAGTCA 3600
QY 4191 GACATGAAGGAGAGATAATGATGCCATACGAAATTTCTAGAAAATGAAAGTAACTTTATA 4250
DB GACATGAAGGAGAGATAATGATGCCATACGAAATTTCTAGAAAATGAAAGTAACTTTATA 3660
QY 4251 GTTACAGAAAGCAATCAGGCGCAGGATAGAGGCTCACACCTGTAATCCCGAGCACTTTGA 4310
DB GTTACAGAAAGCAATCAGGCGCAGGATAGAGGCTCACACCTGTAATCCCGAGCACTTTGA 3720
QY 4311 GAGGCCAGTGGGAAGATTGCTAGAACTCAGGAGTTCAAGAGTTCAGGCTGGGCAACACAGT 4370

Db	3721	GAGGCCACGTGGGAAGATTGCTAGAACTCAGAGTTCAAGACCGCTGGGCAACACAGT	3780
Qy	4371	GAACATCCATTCTCCACAAAAATGGGAAAAAGAAACAAATCAAGTGGTCTCCTGTGG	4430
Db	3781	GAACATCCATTCTCCACAAAAATGGGAAAAAGAAACAAATCAAGTGGTCTCCTGTGG	3840
Qy	4431	GGAGGGAAAGGACTGCAAAAGAGGAAGAGCTCTCGTGGGTGAGGGTGGTGAATTCAGGT	4490
Db	3841	GGAGGGAAAGGACTGCAAAAGAGGAAGAGCTCTCGTGGGTGAGGGTGGTGAATTCAGGT	3900
Qy	4491	TCGTATCTGACTGTGGTAGCAGTTTGGGGTGTATACATCCAAAAATATTCGTAGAATT	4550
Db	3901	TCGTATCTGACTGTGGTAGCAGTTTGGGGTGTATACATCCAAAAATATTCGTAGAATT	3960
Qy	4551	ATGCATCTTAAATGGGTGAGTTTACTGTATGAATATACCTCAATGTAAAGAAAAAT	4610
Db	3961	ATGCATCTTAAATGGGTGAGTTTACTGTATGAATATACCTCAATGTAAAGAAAAAT	4020
Qy	4611	AATGTGTAAAGAAATTTCAATCTCTTCCAGCAAAAGCTTTATTCAAATTCCTGAGCCCT	4670
Db	4021	AATGTGTAAAGAAATTTCAATCTCTTCCAGCAAAAGCTTTATTCAAATTCCTGAGCCCT	4080
Qy	4671	TTACTTCGCAAAATCTCTGCACTTCTGCCCCGTACCATAGGTGACAGCACTAGCTCCAC	4730
Db	4081	TTACTTCGCAAAATCTCTGCACTTCTGCCCCGTACCATAGGTGACAGCACTAGCTCCAC	4140
Qy	4731	AAATTTGATAAATGCATTTCTGAAAAGACTPAGGACAAAAATCAGGCATCATCTGTGCT	4790
Db	4141	AAATTTGATAAATGCATTTCTGAAAAGACTPAGGACAAAAATCAGGCATCATCTGTGCT	4200
Qy	4791	TTCATATCAACACGCTGTACAGCTTGTGTGCTGTCTGCAGCTGCAATGGGACTCTTG	4850
Db	4201	TTCATATCAACACGCTGTACAGCTTGTGTGCTGTCTGCAGCTGCAATGGGACTCTTG	4260
Qy	4851	ATTTCTTTTAAAGAACTTTGGTTTACCAGAGTATTTCCACAAATCTATTCAAAATTTAGTGC	4910
Db	4261	ATTTCTTTTAAAGAACTTTGGTTTACCAGAGTATTTCCACAAATCTATTCAAAATTTAGTGC	4320
Qy	4911	TTATGATATGAACACACTGTGCTAGAGCGCAGAAACAAAGAGGAGGAGAAATCAGTCA	4970
Db	4321	TTATGATATGAACACACTGTGCTAGAGCGCAGAAACAAAGAGGAGGAGAAATCAGTCA	4380
Qy	4971	TTATGTGGGAAACATAGCAAGATTTTATAGATCATTTTCTAGTATTTTAAAGAGCAGCAG	5030
Db	4381	TTATGTGGGAAACATAGCAAGATTTTATAGATCATTTTCTAGTATTTTAAAGAGCAGCAG	4440
Qy	5031	AGTCAAAATCACACATGCAATCAGTATATATCCAAATCATGTAATATGTCCTGTAGAA	5090
Db	4441	AGTCAAAATCACACATGCAATCAGTATATATCCAAATCATGTAATATGTCCTGTAGAA	4500
Qy	5091	AGACTAGAGGAATAAACACAGAAATCTTAACAGTCTATTGTTCATTAGACACTAGTCTAAT	5150
Db	4501	AGACTAGAGGAATAAACACAGAAATCTTAACAGTCTATTGTTCATTAGACACTAGTCTAAT	4560
Qy	5151	TATATATTATTAGACACTATGATATTTGAGATTTTAAAAATCTTTAATATTTAAAAATTTA	5210
Db	4561	TATATATTATTAGACACTATGATATTTGAGATTTTAAAAATCTTTAATATTTAAAAATTTA	4620
Qy	5211	GAGCTCTCTATTTTCCATAGTATTCAGTTTGGCAATGCAATCAAGTATTTACTCTTCTT	5270
Db	4621	GAGCTCTCTATTTTCCATAGTATTCAGTTTGGCAATGCAATCAAGTATTTACTCTTCTT	4680
Qy	5271	TTTTTTTTTTTTTTTTTTTTTTTTTTTGGATGGATTTTGGTCTTGTGGCCATGCTGGAGTG	5330
Db	4681	TTTTTTTTTTTTTTTTTTTTTTTTTTTGGATGGATTTTGGTCTTGTGGCCATGCTGGAGTG	4740
Qy	5331	GAATGGCATGAYCATAGCTCACTGCAACCTCCACCTCTCGGGTTCAGCAAGACTGTGCG	5390
Db	4741	GAATGGCATGAYCATAGCTCACTGCAACCTCCACCTCTCGGGTTCAGCAAGACTGTGCG	4800
Qy	5391	CTCAGCCTCCCGGTAGATGGGATTACAGCGCCGCCACACACACTCGGCTAATCTTGT	5450
Db		TT	

Db	4801	CTCAGCCTCCCGGTAGATGGGATTACAGCGCCGCCACACACACTCGGCTAATCTTGT	4860
Qy	5451	ATTTTTAGTAGAGATGGGTTTCCACCATGTTGGCCAGGCTGTCTCAAACCTCTGACCTC	5510
Db	4861	ATTTTTAGTAGAGATGGGTTTCCACCATGTTGGCCAGGCTGTCTCAAACCTCTGACCTC	4920
Qy	5511	AGAG 5514	
Db	4921	AGAG 4924	
Db		TTTT	

RESULT 2	AC092184	71132 bp	DNA	linear	PRI 12-JUN-2002
LOCUS	AC092184	71132 bp	DNA	linear	PRI 12-JUN-2002
DEFINITION	Homo sapiens 12 BAC RP11-438L7 (Roswell Park Cancer Institute Human BAC Library) complete sequence.				
ACCESSION	AC092184 AC013443				
VERSION	AC092184.7				
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 71132)				
	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chaves,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwater,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,I., Howard,S., Huber,J., Hulyk,S., Hume,J., Joshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Merscher,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Monabhat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,M., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.B., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shostkari,N., Sisson,I., Sodergren,E., Sona				

REFERENCE
AUTHORS
TITLE
JOURNAL

Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 71132)
Worley, K.C.
Direct Submission
Submitted (18-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 71132)
Worley, K.C.
Direct Submission
Submitted (25-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 71132)
Worley, K.C.
Direct Submission
Submitted (12-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 25, 2002 this sequence version replaced gi:20901754.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (12-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (12-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES
source
Location/Qualifiers
1..71132
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-438L7"
complement(1..1448)
/notes="overlaps bases 1..1448 of clone AC092490"
/function="clone overlap"
187..408
/standard_name="57233"
439..560
/standard_name="92005"
774..881
repeat_region
/rpt_family="MIR"

misc_feature
complement(1..1448)
/notes="overlaps bases 1..1448 of clone AC092490"
/function="clone overlap"
187..408
/standard_name="57233"
439..560
/standard_name="92005"
774..881
repeat_region
/rpt_family="MIR"

STS
187..408
/standard_name="57233"
439..560
/standard_name="92005"
774..881
repeat_region
/rpt_family="MIR"

repeat_region
903..1190
/rpt_family="AluSq"
1191..1213
repeat_region
1191..1213
/rpt_family="AT_rich"
1591..1807
STS
/standard_name="6612"
1744..1819
STS
/standard_name="8198"
1966..2264
repeat_region
1966..2264
/rpt_family="AluSq"
3296..3328
repeat_region
3296..3328
/rpt_family="(TTTC)n"
4097..4249
repeat_region
4097..4249
/rpt_family="AluSq"
4360..4654
repeat_region
4360..4654
/rpt_family="AluY"
complement(5140..5262)
repeat_region
complement(5140..5262)
/rpt_family="FLAM_C"
5275..5304
repeat_region
5275..5304
/rpt_family="AT_rich"
5305..5563
repeat_region
5305..5563
/rpt_family="AluJo"
5564..5599
repeat_region
5564..5599
/rpt_family="AT_rich"
5644..5733
repeat_region
5644..5733
/rpt_family="MSTD"
5734..6012
repeat_region
5734..6012
/rpt_family="AluSq"
6020..6336
repeat_region
6020..6336
/rpt_family="AluSx"
6337..6629
repeat_region
6337..6629
/rpt_family="MSTD"
6630..6650
repeat_region
6630..6650
/rpt_family="(CAA)n"
complement(6711..6817)
repeat_region
complement(6711..6817)
/rpt_family="MIR"
complement(7033..7327)
repeat_region
complement(7033..7327)
/rpt_family="AluSx"
7607..7745
repeat_region
7607..7745
/rpt_family="FLAM_A"
7980..8047
repeat_region
7980..8047
/rpt_family="MIR"
complement(8050..8549)
repeat_region
complement(8050..8549)
/rpt_family="LTRA7A"
8567..8698
repeat_region
8567..8698
/rpt_family="MIR"
complement(8815..9101)
repeat_region
complement(8815..9101)
/rpt_family="AluSg"
complement(11227..12175)
repeat_region
complement(11227..12175)
/rpt_family="LTRS"
12625..12900
STS
12625..12900
/standard_name="87432"
complement(13547..13664)
repeat_region
complement(13547..13664)
/rpt_family="LMB5"
complement(13665..13979)
repeat_region
complement(13665..13979)
/rpt_family="AluY"
complement(13980..14193)
repeat_region
complement(13980..14193)
/rpt_family="LMB5"
14622..14924
repeat_region
14622..14924
/rpt_family="AluY"
complement(15813..17136)
repeat_region
complement(15813..17136)
/rpt_family="LIP8"
complement(17272..17395)
repeat_region
complement(17272..17395)
/rpt_family="FLAM_C"
17426..17575
repeat_region
17426..17575
/rpt_family="MIR"
19021..19314
repeat_region
19021..19314
/rpt_family="MIR"

Query Match

Best Local Similarity 84.3%; Score 4649; DB 9; Length 71132;
Matches 5499; Conservative 0; Mismatches 11; Indels 4; Gaps 1;

Qy 1 ACAGACGAATACATGGTCCAAAGCTAGGCTATTGATTTTGAATCATCAAGGTATAGTG 60

Db	33959	ACAGACGATACATGGTCCAAAGCTAGGCGCTATTGATTGAAATCATCAAGGATATAGATG	34018		Db	35039	GGGAGGCAAGAAGACACACTCTGGACACCACCTATGGACAGGTAAAGAGGCGACTCTTCTCGTG	35098
Qy	61	GTATCAAAGGCTTGAGCGAGGAAGAGACGAGACCCCTAGCTGCTTGGCTTAGCATGGCA	120		Qy	1141	GGTGATTGCACTGCGCTTCCCTCTCAGAGCAAAATCTGAGTAATCAGACTGCTAGCTATCC	1200
Db	34019	GTATCAAAGGCTTGAGCGAGGAAGAGACGAGACCCCTAGCTGCTTGGCTTAGCATGGCA	34078		Db	35099	GGTGATTGCACTGCGCTTCCCTCTCAGAGCAAAATCTGAGTAATCAGACTGCTAGCTATCC	35158
Qy	121	TCCTAGCACCTGGCATAGTTCCATTAAACAGTAGGCATGAAGTATCTACTCAGTGAATA	180		Qy	1201	TTTCTCTCATGTAACTGTCTGACTGATGAAGATCAGCTTGTGATCAATATGCAATATATATTTT	1260
Db	34079	TCCTAGCACCTGGCATAGTTCCATTAAACAGTAGGCATGAAGTATCTACTCAGTGAATA	34138		Db	35159	TTTCTCTCATGTAACTGTCTGACTGATGAAGATCAGCTTGTGATCAATATGCAATATATATTTT	35218
Qy	181	AATAGAATGCATATGGCTACAGTACAGAGAGAGAAATAAATCTTTAAATAGACCAAGTTC	240		Qy	1261	TTGATCTGTCTCCTTTTCTTCTTATCAGATCTTATACGCTGTGAGCCCAATTTCTTCTGT	1320
Db	34139	AATAGAATGCATATGGCTACAGTACAGAGAGAGAAATAAATCTTTAAATAGACCAAGTTC	34198		Db	35219	TTGATCTGTCTCCTTTTCTTCTTATCAGATCTTATACGCTGTGAGCCCAATTTCTTCTGT	35278
Qy	241	TATGAGAGCAAAATTAAGTCTTTTATTTGAAGATCTTTAGCTGTGTTTCCAAATTCAG	300		Qy	1321	TTGATCTGTCTCCTTTTCTTCTTATCAGATCTTATACGCTGTGAGCCCAATTTCTTCTGT	1380
Db	34199	TATGAGAGCAAAATTAAGTCTTTTATTTGAAGATCTTTAGCTGTGTTTCCAAATTCAG	34258		Db	35279	TTGATCTGTCTCCTTTTCTTCTTATCAGATCTTATACGCTGTGAGCCCAATTTCTTCTGT	35338
Qy	301	TGCAGCCAGTTAGACACTGATCTCTGCTGGTGAACAAGCATTTTGTATTTGGGGGAC	360		Qy	1381	ACTGATTCGTCCTGAGATTTGTACCATGGTTGAAACTAATTTATGCTAATATATATTAACA	1440
Db	34259	TGCAGCCAGTTAGACACTGATCTCTGCTGGTGAACAAGCATTTTGTATTTGGGGGAC	34318		Db	35339	ACTGATTCGTCCTGAGATTTGTACCATGGTTGAAACTAATTTATGCTAATATATTAACA	35398
Qy	361	TGCTGCTGCTCTGACTCCAAATTAAGGATTTTCTTCTTAAAGACATGGCTCATG	420		Qy	1441	TAGCAAACTCTTAGAGACTCAAAATCATGAAAAGGTAAATAGCAGTACTGTACTAAAAACGG	1500
Db	34319	TGCTGCTGCTCTGACTCCAAATTAAGGATTTTCTTCTTAAAGACATGGCTCATG	34378		Db	35399	TAGCAAACTCTTAGAGACTCAAAATCATGAAAAGGTAAATAGCAGTACTGTACTAAAAACGG	35458
Qy	421	CAAAAATCACTCTTTGGTGTAAATATCTAGTCTTCAAGCAATCTTGTAAATGCAATCAGA	480		Qy	1501	TAGTGTAAATTTTTCGTAATAATTTTCTAAATATTTCAACAGTAAACAACTTGAAGACACA	1560
Db	34379	CAAAAATCACTCTTTGGTGTAAATATCTAGTCTTCAAGCAATCTTGTAAATGCAATCAGA	34438		Db	35459	TAGTGTAAATTTTTCGTAATAATTTTCTAAATATTTCAACAGTAAACAACTTGAAGACACA	35518
Qy	481	AGAAAAAATCCATGTTTGGGAGCAAAATTTTGTCTTAAATCTATATACTGA	540		Qy	1561	CTTTCCTTAGGGAGGCGTTACTGAAATAATTTAGCTATAGTAAGAAAAATTTGTAAATTTAG	1620
Db	34439	AGAAAAAATCCATGTTTGGGAGCAAAATTTTGTCTTAAATCTATATACTGA	34498		Db	35519	CTTTCCTTAGGGAGGCGTTACTGAAATAATTTAGCTATAGTAAGAAAAATTTGTAAATTTAG	35578
Qy	541	GTTCAATTTGCTTAACTGCAAGCAGAGCTGCTAGTGCCTGTCTGACTGAGGTTTCAGAG	600		Qy	1621	AAATGCCAAGCAATCTAAATTTAATTTGCTTGAAGTCACTATGATTTGTGTCCATTTAAGG	1680
Db	34499	GTTCAATTTGCTTAACTGCAAGCAGAGCTGCTAGTGCCTGTCTGACTGAGGTTTCAGAG	34558		Db	35579	AAATGCCAAGCAATCTAAATTTAATTTGCTTGAAGTCACTATGATTTGTGTCCATTTAAGG	35638
Qy	601	AGACTGTGGAAATATGGGGGAATTAGAGGCTATCTGAGGCTCTTCAACAACAATAACCCAA	660		Qy	1681	AGACAAATTCATTCAAGCAAGTTATTAAATGTTAAAGGCCCAATTTGTTAGGCAGTTAATG	1740
Db	34559	AGACTGTGGAAATATGGGGGAATTAGAGGCTATCTGAGGCTCTTCAACAACAATAACCCAA	34618		Db	35639	AGACAAATTCATTCAAGCAAGTTATTAAATGTTAAAGGCCCAATTTGTTAGGCAGTTAATG	35698
Qy	661	GAAGCTATTAAATGCTCTTTAAGGTATTACATAAAATATTACTATCTCATTTGCTTT	720		Qy	1741	GCACATTTTACTATTAACTTAATCTTTCCATTTGTTCCAGACGTAGCTTTAACCCTCTTAG	1800
Db	34619	GAAGCTATTAAATGCTCTTTAAGGTATTACATAAAATATTACTATCTCATTTGCTTT	34678		Db	35699	GCACATTTTACTATTAACTTAATCTTTCCATTTGTTCCAGACGTAGCTTTAACCCTCTTAG	35758
Qy	721	TATTTTGTATTATCATGATTAATTAAGTGTCTACTGTTTACTGCCCTCTGATCTTTGC	780		Qy	1801	GTGTGAATTTGGTTAAGGTCCTCATAAATGCTTTTATGTCAGTCTTTTGTATAGTTTATGTT	1860
Db	34679	TATTTTGTATTATCATGATTAATTAAGTGTCTACTGTTTACTGCCCTCTGATCTTTGC	34738		Db	35759	GTGTGAATTTGGTTAAGGTCCTCATAAATGCTTTTATGTCAGTCTTTTGTATAGTTTATGTT	35818
Qy	781	TAGCTATGGAGCATGGACTGGGCTTTTAGAGCAGCAGCCCAAGGAACCTTAACATTTAA	840		Qy	1861	CATAGAATTTATCTTAATTTCCCTACATTTATGATTTACTATGGATGTATGAGAAATACACCTA	1920
Db	34739	TAGCTATGGAGCATGGACTGGGCTTTTAGAGCAGCAGCCCAAGGAACCTTAACATTTAA	34798		Db	35819	CATAGAATTTATCTTAATTTCCCTACATTTATGATTTACTATGGATGTATGAGAAATACACCTA	35878
Qy	841	AGCAGAGCTGCCCTCAATGGTTTAACTGTGTGACTCTGCCTATGACAGCCCAACCCACC	900		Qy	1921	ATCCTTATACCTTACCTCAATTTTACTCTTTTATAAAGAACTTACATTTACAGAAATAAGA	1980
Db	34799	AGCAGAGCTGCCCTCAATGGTTTAACTGTGTGACTCTGCCTATGACAGCCCAACCCACC	34858		Db	35879	ATCCTTATACCTTACCTCAATTTTAACTCTTTTATAAAGAACTTACATTTACAGAAATAAGA	35938
Qy	901	CATCTTCACTGGATCCCAATCAGGAGCAAGCCGTTGGGTTACCTGTGGGGGTGATGCT	960		Qy	1981	TTTTTTAAAAATATATTTTTTTTGTAGAGACAGGCTCTTTAGCCCAAGCCGAGGCTGTCTCT	2040
Db	34859	CATCTTCACTGGATCCCAATCAGGAGCAAGCCGTTGGGTTACCTGTGGGGGTGATGCT	34918		Db	35939	TTTTTTAAAAATATATTTTTTTTGTAGAGACAGGCTCTTTAGCCCAAGCCGAGGCTGTCTCT	35998
Qy	961	GTACGGGAGGAGCCCAAGGCAAGCTCAAAATTTGAATGTGAAGGSCCAATGCATGCT	1020		Qy	2041	AAATCTCTGGCCCAAGCCGATCTCTGCTGGGCCCTCTTAAAGTGTCTGGAATTTATAGACAT	2100
Db	34919	GTACGGGAGGAGCCCAAGGCAAGCTCAAAATTTGAATGTGAAGGSCCAATGCATGCT	34978		Db	35999	AAATCTCTGGCCCAAGCCGATCTCTGCTGGGCCCTCTTAAAGTGTCTGGAATTTATAGACAT	36058
Qy	1021	CAGACTGAGACAGAACCATCATTAATTAAGTGAGATTTTCTGGCCTGAGACTTGCA	1080		Qy	2101	GAGCCATCACATCCCAATATACAGAAATAAGATTTTTTAAATGAGGATTTTAAATGTTCTTCAG	2160
Db	34979	CAGACTGAGACAGAACCATCATTAATTAAGTGAGATTTTCTGGCCTGAGACTTGCA	35038		Db	36059	GAGCCATCACATCCCAATATACAGAAATAAGATTTTTTAAATGAGGATTTTAAATGTTCTTCAG	36118
Qy	1081	GGGAGGCAAGAGACACTCTCGACACCACTATGACAGGTAAAGGCGAGTCTTCTCTGTG	1140		Qy	2161	AAAAATTTCTTGGAGGTGAGGTCAGACAAATGTCAAAATGTCTCCTCAGTTTACACTGAGATTTTGA	2220
Db					Db	36119	AAAAATTTCTTGGAGGTGAGGTCAGACAAATGTCAAAATGTCTCCTCAGTTTACACTGAGATTTTGA	36178


```

|||||
10 GAACCATCATTAATGAAGTGGAGATTTTCGGCTGAGACTTGCGAGGCGAAGA 69
|||||
Oy 1095 CACTCTGGACACCACTATGGACAG 1118
|||||
Db 70 CACTCTGGACACCACTATGGACAG 93
|||||

RESULT 4
AB040431
LOCUS      2791 bp      mRNA      linear      PRI 03-OCT-2000
DEFINITION Homo sapiens AID mRNA for activation-induced cytidine deaminase,
complete CDS.
ACCESSION  AB040431
VERSION     AB040431.1 GI:9988409
KEYWORDS   AID; activation-induced cytidine deaminase; Human AID.
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
REFERENCE   1 (sites)
AUTHORS     Muto,T., Muramatsu,M., Taniwaki,M., Kinoshita,K. and Honjo,T.
TITLE       Isolation, tissue distribution, and chromosomal localization of the
human activation-induced cytidine deaminase (AID) gene
JOURNAL     Genomics 68 (1), 85-88 (2000)
MEDLINE     20408890
REFERENCE   2 (sites)
AUTHORS     Revy,P., Muto,T., Levy,Y., Geissmann,F., Plebani,A., Sanal,O.,
Catalan,N., Forveille,M., Dufourcq-Lagelouse,R., Gennery,A.,
Tescan,I., Ersoy,F., Kayserili,H., Ugazio,A.G., Brousse,N.,
Muramatsu,M., Notarangelo,L.D., Kinoshita,K., Honjo,T., Fischer,A.
and Durandy,A.
TITLE       Activation-induced cytidine deaminase (AID) deficiency causes the
autosomal recessive form of the Hyper-IgM syndrome (HIGM2)
JOURNAL     Cell 102 (5), 565-575 (2000)
MEDLINE     20460341
REFERENCE   3 (bases 1 to 2791)
AUTHORS     Muto,T., Muramatsu,M., Taniwaki,M., Kinoshita,K. and Honjo,T.
TITLE       Direct Submission
JOURNAL     Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department
of Medical Chemistry, Faculty of Medicine; Yoshida, Sakyo-ku,
Kyoto, Kyoto 606-8501, Japan (E-mail:honjoemfour.med.kyoto-u.ac.jp,
Tel:81-75-753-4371(ex.4371), Fax:81-75-753-4388)
FEATURES
    source      1..2791
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
    gene        1..2791
                /gene="AID"
    CDS         77..673
                /gene="AID"
                /codon_start=1
                /product="activation-induced cytidine deaminase"
                /protein_id="BAB12721.1"
                /db_xref="GI:9988410"
                /translation="MDSLMMNRKFLYQKVRWAKGRRETYLCVVKRRDSATSFSL
                DFGYLRLKNGCHVELFLRYISDMDLDPGRCYRYTWFTSPSCDCAHVADFLGNP
                NLSRIPTARLYFQEDKRAEPEGLRLHRAGVQIAIMTFKDYFCWNTFVENHETFK
                AWEGHNSVRLSRLRILLPLLYEVDLDRDAFRTLGI"
BASE COUNT    842 a 548 c 625 g 776 t
ORIGIN

Query Match      1.5%; Score 84; DB 9; Length 2791;
Best Local Similarity 100.08; Pred. No. 6.9e-34;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1035 GAACCATCATTAATGAAGTGGAGATTTTCGGCTGAGACTTGCGAGGCGAAGA 1094
|||||
Db 1 GAACCATCATTAATGAAGTGGAGATTTTCGGCTGAGACTTGCGAGGCGAAGA 60
|||||

Oy 1095 CACTCTGGACACCACTATGGACAG 1118
|||||
Db 61 CACTCTGGACACCACTATGGACAG 84
|||||

```

```

RESULT 5
AL353650/c
LOCUS      84778 bp      DNA      linear      PRI 19-SEP-2001
DEFINITION Human DNA sequence from clone RP11-344N19 on chromosome 10.
Contains STSS, GSSs and a CpG island, complete sequence.
ACCESSION  AL353650
VERSION     AL353650.5 GI:14141281
KEYWORDS   HTG; CpG island.
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 84778)
AUTHORS     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE       Griffiths,C.
JOURNAL     Direct Submission
COMMENT     Submitted (01-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On May 17, 2001 this sequence version replaced gi:13567986.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TrEMBL; Wp:, WormPEP; Information on the WormPEP
database can be found at
http://www.sanger.ac.uk/projects/Celegans/wormpep
This sequence was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-344N19 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-344N19. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP11-344N19 is at 1 in this sequence.
The true left end of clone RP11-735G18 is at 84679 in this
sequence. The true right end of clone RP11-426A21 is at 3263 in
this sequence.
FEATURES
    source      1..84778
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="10"
                /clone="RP11-344N19"
                /clone_lib="RPCI-11.2"
    misc_feature 14..616
                /note="match: GSS: Em:AQ542535"
    repeat_region 102..282
                /note="L1PA7 repeat: matches 5960. .6141 of consensus"
    repeat_region 1753..2298
                /note="L1PA14 repeat: matches 5601. .6149 of consensus"
    repeat_region 4213..4511
                /note="AluX repeat: matches 1. .300 of consensus"
    repeat_region 4643..4720
                /note="MIR repeat: matches 168. .245 of consensus"
    repeat_region 4733..5068
                /note="MER2 repeat: matches 8. .345 of consensus"

```



```

AC004971/c
LOCUS       AC004971      123253 bp      DNA      linear      PRI 30-SEP-2000
DEFINITION  Homo sapiens PAC clone RP5-1125K23 from 7, complete sequence.
ACCESSION   AC004971
VERSION     AC004971.3  GI:10047915
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens.
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1  (bases 1 to 123253)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
  PUBMED    9847074
REFERENCE   2  (bases 1 to 123253)
  AUTHORS   Stromatt,C., Ali,J. and Miller,K.
  TITLE     The sequence of Homo sapiens PAC clone RP5-1125K23
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 123253)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
            Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
  JOURNAL   MO 63108, USA
REFERENCE   4  (bases 1 to 123253)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
            Submitted (10-SEP-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
  JOURNAL   MO 63108, USA
REFERENCE   5  (bases 1 to 123253)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
            Submitted (30-SEP-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Sep 10, 2000 this sequence version replaced gi:7630775.
  JOURNAL   ----- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            ----- Summary Statistics
            -----
            Center project name: H_DJ1125K23
            -----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/CTB/CHR7 , send
mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-5, prepared by
Pieter de Jong and coworkers at the Roswell Park Cancer Institute

```

(http://bacpac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.

VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP4-607J2; the clone sequenced to the right is RP5-826E18, 200 base pair overlap. Actual end of this clone is at base position 123057 of RP5-1125K23.

There are polymorphic base pair differences in the overlap between the clone RP5-1125K23 and RP4-607J2.

FEATURES

source	Location/Qualifiers
1..123253	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7"
	/clone="RP5-1125K23"
	/clone_lib="RPCI-5"
65..226	
misc_feature	/note="similar to EST AA158750 (NID:g1733597) zo63e09.r1"
65..225	
misc_feature	/note="similar to EST AA079742 (NID:g1618634) zm20g10.r1"
65..225	
misc_feature	/note="similar to EST AA486040 (NID:g2216256) ab40c06.r1"
65..225	
misc_feature	/note="similar to EST AW579567 (NID:g7254616)"
65..225	
misc_feature	/note="similar to EST BE146164 (NID:g8608888)"
65..210	
misc_feature	/note="similar to EST AA679866 (NID:g2656333) ac83h04.s1"
68..225	
misc_feature	/note="similar to EST AW804699 (NID:g7856569)"
68..225	
misc_feature	/note="similar to EST AW900299 (NID:g8064504)"
77..225	
misc_feature	/note="similar to EST AW320240 (NID:g6749784) uo16c08.y1"
325..616	
repeat_region	/rpt_family="Alu"
837..965	
misc_feature	/note="similar to EST AA079742 (NID:g1618634) zm20g10.r1"
837..965	
misc_feature	/note="similar to EST AW320240 (NID:g6749784) uo16c08.y1"
837..965	
misc_feature	/note="similar to EST AW579567 (NID:g7254616)"
837..965	
misc_feature	/note="similar to EST AW900299 (NID:g8064504)"
837..965	
misc_feature	/note="similar to EST BE146164 (NID:g8608888)"
837..959	
misc_feature	/note="similar to EST AW804699 (NID:g7856569)"
837..954	
misc_feature	/note="similar to EST AA486040 (NID:g2216256) ab40c06.r1"
865..965	
misc_feature	/note="similar to EST AA914792 (NID:g3054184) vz09c01.r1"
1113..1425	
repeat_region	/rpt_family="Alu"
1470..1527	
misc_feature	/note="similar to EST AA914792 (NID:g3054184) vz09c01.r1"
1470..1527	
misc_feature	/note="similar to EST AW320240 (NID:g6749784) uo16c08.y1"
1470..1527	
misc_feature	/note="similar to EST AW579567 (NID:g7254616)"
1470..1527	
misc_feature	/note="similar to EST BE146164 (NID:g8608888)"
1470..1521	
misc_feature	/note="similar to EST AA079742 (NID:g1618634) zm20g10.r1"
1470..1509	
misc_feature	/note="similar to EST AW900299 (NID:g8064504)"
1831..1952	
misc_feature	

misc_feature /note="similar to EST AA233940 (NID:g1855135) zr27a04.r1" 1831..1952

misc_feature /note="similar to EST AW674791 (NID:g7540101) ba59f07.y1" 1833..1953

misc_feature /note="similar to EST AA914792 (NID:g3054184) vz09c01.r1" 1833..1952

misc_feature /note="similar to EST AW320240 (NID:g6749784) uol6c08.y1" 1833..1934

misc_feature /note="similar to EST BE146164 (NID:g8608888)" 1833..1925

repeat_region /note="similar to EST AW579567 (NID:g7254616)" 2205..2327

repeat_region /rpt_family="LI" 2338..2644

repeat_region /rpt_family="Alu" 2656..2822

repeat_region /rpt_family="Alu" 2879..3182

repeat_region /rpt_family="Alu" 3317..3425

misc_feature /rpt_family="Alu" 3695..3887

misc_feature /note="similar to EST AW674791 (NID:g7540101) ba59f07.y1" 3695..3866

misc_feature /note="similar to EST AA233940 (NID:g1855135) zr27a04.r1" 3695..3784

misc_feature /note="similar to EST AW320240 (NID:g6749784) uol6c08.y1" 3739..3887

misc_feature /note="similar to EST AI199495 (NID:g3752101) qi56f08.x1" 3862..3887

repeat_region /note="similar to EST AW732308 (NID:g7632637) bb02e08.y1" 4752..5441

repeat_region /rpt_family="ERV1" 5526..5656

repeat_region /rpt_family="Alu" 5769..6067

repeat_region /rpt_family="Alu" 6154..6190

repeat_region /rpt_family="Alu" 6220..6500

repeat_region /rpt_family="Alu" 7622..7687

repeat_region /rpt_family="MIR" 7779..7961

repeat_region /rpt_family="Alu" 7965..8265

misc_feature /rpt_family="Alu" 9058..10603

misc_feature /note="CpG island (%GC=66.6, o/e=1.00, #CpGs=168)" 9177..9758

misc_feature /note="similar to EST AW732308 (NID:g7632637) bb02e08.y1" 9177..9365

misc_feature /note="similar to EST AW674791 (NID:g7540101) ba59f07.y1" 9177..9363

misc_feature /note="similar to EST AI199495 (NID:g3752101) qi56f08.x1" 9177..9363

Query Match 1.4%; Score 75; DB 9; Length 123253;

Best Local Similarity 100.0%; Pred. No. 4.5e-29;

Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5438 GCCTAATGTTGTTATTTAGTAGAGTGGGTTTACCAGTTGGCCAGGCTGCTCA 5497

DB 104338 GCCTAATGTTGTTATTTAGTAGAGTGGGTTTACCAGTTGGCCAGGCTGCTCA 104279

QY 5498 AACTCCTGACCTCAG 5512

DB 104278 AACTCCTGACCTCAG 104264

RESULT 7

AL356418 168524 bp DNA linear HTG 11-JUL-2001

LOCUS AL356418

DEFINITION Homo sapiens chromosome 10 clone RP11-215117, *** SEQUENCING IN

ACCESSION AL356418

VERSION GI:13620377

KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Plumb,B.

TITLE Direct Submission

JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk requests: clonerequest@sanger.ac.uk

COMMENT On Apr 12, 2001 this sequence version replaced gi:9797524.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BA215117

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: piasmld; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 161402 bases at least Q40

Consensus quality: 164147 bases at least Q30

Consensus quality: 165883 bases at least Q20

Insert size: 167324; sum-of-contigs

Insert size: 158470; 5.5% error; agarose-fp

Quality coverage: 4.52x in Q20 bases; sum-of-contigs Quality coverage: 4.85x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 19496: contig of 19496 bp in length

19497 19596: gap of 100 bp

19597 38508: contig of 18912 bp in length

38509 38608: gap of 100 bp

38609 55774: contig of 17166 bp in length

55775 55874: gap of 100 bp

55875 81803: contig of 25929 bp in length

81804 81903: gap of 100 bp

81904 89760: contig of 7857 bp in length

89761 89860: gap of 100 bp

89861 96400: contig of 6540 bp in length

96401 96500: gap of 100 bp

96501 126138: contig of 29638 bp in length

126139 126238: gap of 100 bp

126239 138525: contig of 12287 bp in length

138526 138625: gap of 100 bp

138626 151528: contig of 12903 bp in length

151529 151628: gap of 100 bp

151629 156089: contig of 4461 bp in length

156090 156189: gap of 100 bp

156190 158536: contig of 2347 bp in length

158537 158636: gap of 100 bp

158637 166329: contig of 7693 bp in length

166330 166429: gap of 100 bp

166430 168524: contig of 2095 bp in length.

Location/Qualifiers

1..168524

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="10"

/clone="RP11-215117"

FEATURES

source


```
misc_feature
1. .19496
/note="assembly_fragment:00715
fragment_chain:1"
19597..38508
/note="assembly_fragment:00847
fragment_chain:1"
38609..55774
/note="assembly_fragment:00141
fragment_chain:1"
55875..81803
/note="assembly_fragment:01994
fragment_chain:2"
81904..89760
/note="assembly_fragment:00862
fragment_chain:2"
89861..96400
/note="assembly_fragment:00801
fragment_chain:2"
96501..126138
/note="assembly_fragment:00381
fragment_chain:3"
126239..138525
/note="assembly_fragment:00176
fragment_chain:3"
138626..151528
/note="assembly_fragment:00614
fragment_chain:4"
151629..156089
/note="assembly_fragment:01732
fragment_chain:4"
156190..158536
/note="assembly_fragment:00877"
158637..166329
/note="assembly_fragment:01428"
168430..168524
/note="assembly_fragment:01755"
53308 a 33132 c 33014 g 47867 t 1203 others
ORIGIN
Query Match 1.48; Score 75; DB 2; Length 168524;
Best Local Similarity 100.0%; Pred. No. 4.4e-29;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 5438 GCCTAATGTTGTATTTTGTAGATGGGTTTCACCATGTTGGCAGCTGCTCTCA 5497
|||||
Db 15343 GCCTAATGTTGTATTTTGTAGATGGGTTTCACCATGTTGGCAGCTGCTCTCA 15402
|||||
Qy 5498 AACTCCTGACCTCAG 5512
|||||
Db 15403 AACTCCTGACCTCAG 15417
|||||

RESULT 8
AP001083
LOCUS Homo sapiens chromosome 11 clone CMB9-9E9 map 11q22, WORKING DRAFT
DEFINITION Homo sapiens chromosome 11 clone CMB9-9E9 map 11q22, WORKING DRAFT
SEQUENCE, 5 unordered pieces.
ACCESSION AP001083
VERSION AP001083.2 GI:8117230
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:CMB9-9E9.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 29359)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 29,359 genomic DNA of 11q22
JOURNAL Published Only in DataBase (2000)
REFERENCE 2 (bases 1 to 29359)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
```

```
TITLE
JOURNAL Direct Submission
Submitted (25-JAN-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@sc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
COMMENT
On May 30, 2000 this sequence version replaced gi:6997374.
----- Genomic Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@sc.riken.go.jp
----- Project Information
Center project name: HumDraft11
Center clone name: CMB9-9E9
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 27370 bases at least Q40
Consensus quality: 28188 bases at least Q30
Consensus quality: 28581 bases at least Q20
Insert size: 28959; sum-of-contigs
Quality coverage: 10.68x in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently consists of
5 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 19044 contig of 19044 bp in length
19145 19145 contig of 3467 bp in length
22712 26518 contig of 3807 bp in length
26619 28009 contig of 1391 bp in length
28110 29359 contig of 1250 bp in length
Sequence updated (26-May-2000).
* NOTE: this is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 19044: contig of 19044 bp in length
* 19045 19144: gap of 100 bp
* 19145 22611: contig of 3467 bp in length
* 22612 22711: gap of 100 bp
* 22712 26518: contig of 3807 bp in length
* 26519 26618: gap of 100 bp
* 26619 28009: contig of 1391 bp in length
* 28010 28109: gap of 100 bp
* 28110 29359: contig of 1250 bp in length.
FEATURES
Location/Qualifiers
1..29359
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q22"
/clone="CMB9-9E9"
1..19044
/note="assembly_fragment"
19145..22611
/note="assembly_fragment"
22712..26518
/note="assembly_fragment"
26619..28009
/note="assembly_fragment"
28110..29359
/note="assembly_fragment"
misc_feature
/note="assembly_fragment"
19145..22611
misc_feature
/note="assembly_fragment"
22712..26518
misc_feature
/note="assembly_fragment"
26619..28009
misc_feature
/note="assembly_fragment"
28110..29359
```

```

BASE COUNT      7689 a   6357 c   6443 g   8470 t   400 others
ORIGIN
Query Match
Best Local Similarity 1.3%; Score 74; DB 2; Length 29359;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5438 GGCATATGTTGTATTTAGTAGAGATGGGGTTTCACCATGTTGCCAGGCTGGTCTCA 5497
      |||||||
Db 22056 GGCATATGTTGTATTTAGTAGAGATGGGGTTTCACCATGTTGCCAGGCTGGTCTCA 22115
      |||||||
QY 5498 AACTCTCTGACCTCA 5511
      |||||||
Db 22116 AACTCTCTGACCTCA 22129
      |||||||

RESULT 9
AC024709
LOCUS
DEFINITION Homo sapiens chromosome 3 clone RP11-147G7, WORKING DRAFT SEQUENCE,
19 unordered pieces.
AC024709
VERSION AC024709.5 GI:9838092
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 168079)
Waterston,R.H.
Direct Submission
Submitted (01-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Aug 17, 2000 this sequence version replaced gi:7630865.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0147G07
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 96% of reads
Chemistry: Dye-terminator Big Dye; 4% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 158772 bases at least Q40
Consensus quality: 160752 bases at least Q30
Consensus quality: 162013 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 167609; sum-of-contigs
Quality coverage: 4.90 in Q20 bases; agarose-fp
Quality coverage: 5.01 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1412: contig of 1412 bp in length
* 1413 1512: gap of unknown length
* 1513 3446: contig of 1934 bp in length
* 3447 5274: gap of unknown length
* 5274: contig of 1728 bp in length

```

FEATURES

Location/Qualifiers

Source

```

1..168079
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-147G7"
1..1412
/note="assembly_name:Contig6"
1513..3446
/note="assembly_name:Contig7"
3547..5274
/note="assembly_name:Contig8"
5375..6854
/note="assembly_name:Contig9"
6955..8464
/note="assembly_name:Contig10"
8565..10233
/note="assembly_name:Contig11"
10334..11461
/note="assembly_name:Contig12"
clone_end:SP6
vector_side:right
11562..15323
/note="assembly_name:Contig13"
15424..18905
/note="assembly_name:Contig14"
19006..23670
/note="assembly_name:Contig15"
23771..30058
/note="assembly_name:Contig16"
30159..34404
/note="assembly_name:Contig17"
34505..43175
/note="assembly_name:Contig18"
43276..53770
/note="assembly_name:Contig19"
53871..64674
/note="assembly_name:Contig20"
64775..82701
/note="assembly_name:Contig21"
82802..107645

```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

```
/note="assembly_name:Contig22"
107746. .135351
/note="assembly_name:Contig23"
135452. 168079
/note="assembly_name:Contig24"
BASE COUNT 49957 a 32303 c 31709 g 52284 t 1826 others
ORIGIN

Query Match      1.3%; Score 74; DB 2; Length 168079;
Best Local Similarity 100.0%; Pred. No. 1.5e-28;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5438 GCCTAATCTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGCTCA 5497
      |||
Db 122890 GCTAATCTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGCTCA 122949
      |||

QY 5498 AACTCCTGACCTCA 5511
      |||
Db 122950 AACTCCTGACCTCA 122963
      |||

RESULT 10
AL353749
LOCUS Homo sapiens chromosome 9 clone RP11-138018, *** SEQUENCING IN
DEFINITION PROGRESS ***, 26 unordered pieces.
ACCESSION AL353749
VERSION AL353749.3 GI:9863661
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Plumb,B.
            1 (bases 1 to 172376)
Direct Submission
Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 21, 2000 this sequence version replaced gi:9213188.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba138018
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 159786 bases at least Q40
Consensus quality: 163877 bases at least Q30
Consensus quality: 166550 bases at least Q20
Insert size: 169876; sum-of-contigs
Insert size: 191077; agarose-fp
Quality coverage: 3.70x in Q20 bases; sum-of-contigs Quality
coverage: 3.43x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 17327: contig of 17327 bp in length
* 17328 17427: gap of 100 bp
* 17428 20692: contig of 3265 bp in length
* 20693 20792: gap of 100 bp
* 20793 23652: contig of 2860 bp in length
* 23653 23752: gap of 100 bp
*

misc_feature
/note="assembly_name:Contig22"
25971: contig of 2219 bp in length
25972 26071: gap of 100 bp
26072 46738: contig of 20667 bp in length
46739 46838: gap of 100 bp
46839 56508: contig of 9670 bp in length
56509 56608: gap of 100 bp
56609 59070: contig of 2462 bp in length
59071 59170: gap of 100 bp
59171 61789: contig of 2619 bp in length
61790 61889: gap of 100 bp
61890 72802: contig of 10913 bp in length
72803 72902: gap of 100 bp
72903 75383: contig of 2481 bp in length
75384 75483: gap of 100 bp
75484 78512: contig of 3029 bp in length
78513 78612: gap of 100 bp
78613 80836: contig of 2224 bp in length
80837 80936: gap of 100 bp
80937 88055: contig of 7119 bp in length
88056 88155: gap of 100 bp
88156 92223: contig of 4068 bp in length
92224 92323: gap of 100 bp
92324 96342: contig of 4219 bp in length
96343 96442: gap of 100 bp
96443 99878: contig of 3236 bp in length
99879 99978: gap of 100 bp
99979 105558: contig of 5580 bp in length
105559 105658: gap of 100 bp
105659 122337: contig of 16679 bp in length
122338 122437: gap of 100 bp
12438 128755: contig of 6318 bp in length
128756 128855: gap of 100 bp
128856 131775: contig of 2920 bp in length
131776 131875: gap of 100 bp
131876 136653: contig of 4778 bp in length
136654 136753: gap of 100 bp
136754 149239: contig of 12486 bp in length
149240 149339: gap of 100 bp
149340 156036: contig of 6697 bp in length
156037 156136: gap of 100 bp
156137 161800: contig of 5664 bp in length
161801 161900: gap of 100 bp
161901 168950: contig of 7050 bp in length
168951 169050: gap of 100 bp
169051 172376: contig of 3326 bp in length.

FEATURES
            source
            i. 172376
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="9"
            /clone="RP11-138018"
            /clone_lib="RPC1-11.1"
            i. 17327
            /note="assembly_fragment:00680
            clone_end:sp6
            vector_side:left"
            17428..20692
            /note="assembly_fragment:00043
            fragment_chain:1"
            20793..23652
            /note="assembly_fragment:00840
            fragment_chain:1"
            23753..25971
            /note="assembly_fragment:00060
            fragment_chain:2"
            26072..46738
            /note="assembly_fragment:00750
            fragment_chain:2"
            46839..56508
            /note="assembly_fragment:00077
            fragment_chain:3"
            56609..59070
            /note="assembly_fragment:01416
```

misc_feature
fragment_chain:3"
59171..61789
/note="assembly_fragment:00144
fragment_chain:4"
61890..72802
/note="assembly_fragment:00688
fragment_chain:3"
72903..75383
/note="assembly_fragment:00093"
75484..78512
/note="assembly_fragment:00116"
78613..80836
/note="assembly_fragment:00141"
80937..88055
/note="assembly_fragment:00244"
88156..92223
/note="assembly_fragment:00457"
92324..96342
/note="assembly_fragment:01057"
96643..99878
/note="assembly_fragment:01059"
99979..105558
/note="assembly_fragment:01199"
105659..122337
/note="assembly_fragment:01269"
122438..128755
/note="assembly_fragment:01299"
128856..131775
/note="assembly_fragment:01418"
131876..136653
/note="assembly_fragment:01559"
136754..149239
/note="assembly_fragment:01640"
149340..156036
/note="assembly_fragment:01302
fragment_chain:5"
156137..161800
/note="assembly_fragment:00235
fragment_chain:5"
161901..168950
/note="assembly_fragment:00728
fragment_chain:5"
169051..172376
/note="assembly_fragment:00685
fragment_chain:5
clone_end:T7
vector_side:right"
BASE COUNT 47584 a 38065 c 37064 g 47145 t 2518 others
ORIGIN

Query Match 1.3%; Score 74; DB 2; Length 172376;
Best Local Similarity 100.0%; Pred. No. 1.5e-28;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5438 GCCTAATGTTGTATTATTTAGTAGAGTGGGTTTCACCATGTTGGCCAGGCTGGTCFCA 5497
|||||
Db 14043 GCCTAATGTTGTATTATTTAGTAGAGTGGGTTTCACCATGTTGGCCAGGCTGGTCFCA 14102
|||||
QY 5498 AACTCCTGACCTCA 5511
|||||
Db 14103 AACTCCTGACCTCA 14116
|||||

RESULT 11
AL139113
LOCUS AL139113 193280 bp DNA linear PRI 12-MAY-2001
DEFINITION Human DNA sequence from clone RP11-176F3 on chromosome 9, complete
sequence.
ACCESSION AL139113
VERSION AL139113.21 GI:14041744
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 193280)
Tracey,A.
Direct Submission
Submitted (12-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On May 15, 2001 this sequence version replaced gi:13872249.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em; EMBL; Sw;
SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C-elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-176F3 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
This sequence is the entire insert of clone RP11-176F3 The true
left end of clone RP11-138018 is at 125710 in this sequence. The
true right end of clone RP11-133022 is at 92565 in this sequence.
FEATURES
Location/Qualifiers
1..193280
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-176F3"
/clone_lib="RPCI-11.1"
359..530
/note="MER5B repeat: matches 16..178 of consensus"
545..689
/note="MERSA repeat: matches 34..182 of consensus"
1993..2148
/note="3 copies 52 mer 92% conserved"
4526..4577
/note="26 copies 2 mer aa 80% conserved"
4880..5119
/note="AluSx repeat: matches 1..259 of consensus"
5145..5843
/note="L1LPA16 repeat: matches 5460..6157 of consensus"
7725..7802
/note="MIR repeat: matches 64..150 of consensus"
7860..8316
/note="L1MD repeat: matches 1189..1689 of consensus"
8317..8618
/note="AluSx repeat: matches 1..302 of consensus"
8619..8714
/note="L1MD repeat: matches 1097..1189 of consensus"
8724..9205
/note="L1MC1 repeat: matches 5852..6332 of consensus"
9213..9358
/note="FLAM_C repeat: matches 1..142 of consensus"
9380..9631
/note="AluJo repeat: matches 43..282 of consensus"
9632..9804
/note="L1MC1 repeat: matches 5649..5823 of consensus"

repeat_region 9802..9920
/note="LIMB2 repeat: matches 6005..6114 of consensus"
repeat_region 10339..10370
/note="16 copies 2 mer gt 100% conserved"
repeat_region 10494..10642
/note="MIR repeat: matches 84..254 of consensus"
repeat_region 10818..11127
/note="LIM4 repeat: matches 11..302 of consensus"
repeat_region 11140..11451
/note="AluSx repeat: matches 1..312 of consensus"
repeat_region 11635..11777
/note="LIM4 repeat: matches 470..609 of consensus"
repeat_region 12367..12533
/note="L2 repeat: matches 2362..2530 of consensus"
repeat_region 12662..12914
/note="L2 repeat: matches 1931..2203 of consensus"
repeat_region 12918..13288
/note="LIP1A16 repeat: matches 4675..5036 of consensus"
repeat_region 13289..13579
/note="AluSg repeat: matches 5..295 of consensus"
repeat_region 13580..14394
/note="LIP1A16 repeat: matches 5036..5867 of consensus"
repeat_region 14433..14644
/note="LFR12 repeat: matches 450..671 of consensus"
repeat_region 14647..14780
/note="AluSp/q repeat: matches 159..293 of consensus"
repeat_region 14782..15061
/note="AluSg repeat: matches 33..299 of consensus"
repeat_region 15064..16072
/note="L1 repeat: matches 3303..4289 of consensus"
repeat_region 16073..16371
/note="AluSx repeat: matches 1..299 of consensus"
repeat_region 16372..16513
/note="L1 repeat: matches 4289..4432 of consensus"
repeat_region 16525..16632
/note="AluSg/x repeat: matches 1..108 of consensus"
repeat_region 16721..17032
/note="AluSx repeat: matches 1..308 of consensus"
repeat_region 17247..17299
/note="L2 repeat: matches 2699..2750 of consensus"
repeat_region 17451..17521
/note="L2 repeat: matches 2624..2694 of consensus"
repeat_region 18163..18761
/note="LIMC3 repeat: matches 7110..7727 of consensus"
repeat_region 18766..19053
/note="AluJb repeat: matches 1..289 of consensus"
repeat_region 19115..19217
/note="MIR repeat: matches 132..248 of consensus"
repeat_region 19230..19404
/note="AluSg/x repeat: matches 134..308 of consensus"
misc_feature 20104..21605
/note="CpG Island"
/evidence=not_experimental
20420..20745
/note="163 copies 2 mer cc 57% conserved"
21669..21702
/note="17 copies 2 mer gg 85% conserved"
21833..21854
/note="11 copies 2 mer aa 100% conserved"
22944..23239
/note="AluSg repeat: matches 1..296 of consensus"
24478..24760
/note="AluJc repeat: matches 1..278 of consensus"
24924..25224
/note="AluSx repeat: matches 1..296 of consensus"
25400..25694
/note="AluSp repeat: matches 1..295 of consensus"
25730..26027
/note="AluJc repeat: matches 1..298 of consensus"
26192..26508
/note="AluSx repeat: matches 1..312 of consensus"
26509..26640
/note="FLAM_C repeat: matches 1..135 of consensus"

repeat_region 26684..26978
/note="AluSx repeat: matches 1..295 of consensus"
repeat_region 27513..27657
/note="AluSx repeat: matches 1..150 of consensus"
repeat_region 27658..27957
/note="AluY repeat: matches 1..297 of consensus"
repeat_region 27958..28125
/note="AluSx repeat: matches 150..302 of consensus"
repeat_region 28467..28505
/note="AluY repeat: matches 1..47 of consensus"
repeat_region 28548..28843
/note="AluSx repeat: matches 1..301 of consensus"
repeat_region 29177..29465
/note="AluSx repeat: matches 15..299 of consensus"
repeat_region 29720..30029
/note="AluY repeat: matches 1..296 of consensus"
repeat_region 30103..30417
/note="AluSx repeat: matches 5..297 of consensus"
repeat_region 30473..30639
/note="AluJb repeat: matches 129..291 of consensus"
repeat_region 30662..30954
/note="AluSg repeat: matches 1..293 of consensus"
repeat_region 30992..31302
/note="AluSx repeat: matches 1..311 of consensus"
repeat_region 31327..31620
/note="AluSx repeat: matches 1..298 of consensus"
repeat_region 31698..31856
/note="AluSg/x repeat: matches 131..298 of consensus"
repeat_region 31857..32040
/note="AluSx repeat: matches 137..308 of consensus"
repeat_region 32041..32337
/note="AluSg repeat: matches 1..296 of consensus"
repeat_region 32338..32470
/note="AluSx repeat: matches 1..137 of consensus"
repeat_region 32668..32848
/note="LIPB3 repeat: matches 5974..6146 of consensus"
repeat_region 33097..33399
/note="AluSp repeat: matches 1..304 of consensus"
repeat_region 33462..33759
/note="AluSx repeat: matches 1..299 of consensus"
repeat_region 33792..34088
/note="AluSx repeat: matches 5..301 of consensus"
repeat_region 34281..34377
/note="HX3 repeat: matches 1..97 of consensus"
repeat_region 34378..34421
/note="22 copies 2 mer aa 79% conserved"

Query Match 1.3%; Score 74; DB 9; Length 193280;
Best Local Similarity 100.0%; Pred. No. 1.5e-28;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5438 GGCTAATGTTGTTATTTTAGTAGACATGGGGTTTACCATGTTGGCAGGCTGCTCA 5497
|||||
Db 139758 GGCTAATGTTGTTATTTTAGTAGAGATGGGGTTTACCATGTTGGCAGGCTGCTCA 139817
|||||
QY 5498 AACTCCTGACCTCA 5511
|||||
Db 139818 AACTCCTGACCTCA 139831
|||||

RESULT 12
AC048334/C
LOCUS 215719 bp DNA linear PRI 04-DEC-2001
DEFINITION Homo sapiens 3 BAC RP11-572C15 (Roswell Park Cancer Institute Human
BAC library) complete sequence.
ACCESSION AC048334
VERSION AC048334.21 GI:15027693
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 215719)

AUTHORS

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbaria, J., Benton, J., Blimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, I.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., He, X., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Honsl, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louissegh, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Moore, S., Morgan, M., Moorish, T., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Naylor, S.L., Weinstein, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 215719)
Worley, K.C.

Direct Submission
Submitted (14-APR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 215719)
Worley, K.C.

Direct Submission
Submitted (01-AUG-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 215719)
Worley, K.C.

Direct Submission
Submitted (27-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 215719)
Worley, K.C.

Direct Submission
Submitted (04-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 28, 2001 this sequence version replaced gi:15022627.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of

the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.

Genes and region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES	Source	Location/Qualifiers
misc_feature		1..215719 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="3" /clone="RP11-572C15" 1..2296 /note="overlaps bases 3705..6000 of clone AC069516"
repeat_region		/function="clone overlap" complement(447..556) /rpt_family="L2"
repeat_region		623..1065 /rpt_family="L2"
repeat_region		complement(1882..2179) /rpt_family="AluSq"
repeat_region		3667..3843 /rpt_family="LIME"
repeat_region		complement(3931..4070) /rpt_family="MIR"
repeat_region		complement(4624..4710) /rpt_family="MIR"
repeat_region		4824..5131 /rpt_family="AluSg"
repeat_region		5248..5271 /rpt_family="AT-rich"
repeat_region		complement(5459..5772) /rpt_family="AluSq"
repeat_region		6767..6846 /rpt_family="MER33"
repeat_region		complement(6875..6937) /rpt_family="MER34B"
repeat_region		7090..7350 /rpt_family="MER33"
repeat_region		7371..7529 /rpt_family="MIR"
repeat_region		complement(7818..8112) /rpt_family="AluSq"
repeat_region		complement(8126..8364) /rpt_family="LIME3"
repeat_region		complement(8365..9004) /rpt_family="Ltr8"

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 153468: contig of 153468 bp in length
* 153469 153568: gap of 100 bp
* 153569 153791: contig of 223 bp in length
* 153792 153891: gap of 100 bp
* 153892 173943: contig of 20052 bp in length.
Location/Qualifiers
1. .173943
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosomes="8"
/map="8"
/clone="RP11-139K6"
/clone_lib="RPCI-11 Human Male BAC"
1. 153468
/note="assembly_fragment
clone_end:SP6
vector_side:left"
153569. .153791
/note="assembly_fragment"
153892. 173943
/note="assembly_fragment
clone_end:T7
vector_side:right"
a 35630 c 36092 g 53491 t 202 others
BASE COUNT 48528 a 35630 c 36092 g 53491 t 202 others
ORIGIN
Query Match 1.3%; Score 73; DB 2: Length 173943;
Best Local Similarity 100.0%; Pred. NO. 5.3e-28;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5439 GCTATGTTGTTATTTTAGTAGAGATGGGTTTCCACCATGTGGCCAGGCTGGTCTCAA 5498
|||||
Db 149908 GCTATGTTGTTATTTTAGTAGAGATGGGTTTCCACCATGTGGCCAGGCTGGTCTCAA 149849
|||||
QY 5499 ACTCTGACCTCA 5511
|||||
Db 149848 ACTCTGACCTCA 149836
.
.
RESULT 14
AC027595
LOCUS
DEFINITION Homo sapiens chromosome 8 clone RP11-359A12 map 8, WORKING DRAFT
SEQUENCE, 8 unordered pieces.
ACCESSION AC027595
VERSION AC027595.3 GI:14626930
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 195661)
Birren.B., Linton.L., Nusbaum.C. and Lander.E.
Homo sapiens chromosome 8, clone RP11-359A12
Unpublished
2 (bases 1 to 195661)
Birren.B., Linton.L., Nusbaum.C., Lander.E., Abraham.H., Allen.N.,
Anderson.S., Baldwin.J., Barna.N., Bastien.V., Bedalov.F.,
Boguslavsky.L., Boukhgalter.B., Brown.A., Burkett.G.,
Campoliano.A., Castie.A., Choepel.Y., Colangelo.M., Collins.S.,
Collumore.A., Cooke.P., DeArellano.K., Dewar.K., Diaz.J.S.,
Dodge.S., Domino.M., Doyle.M., Ferreira.P., FitzHugh.W., Gage.D.,
Galagan.J., Gardyna.S., Ginde.S., Goyette.M., Graham.L.,
Grand-Pierre.N., Grant.G., Hagos.B., Heaford.A., Horton.L.,
Howland.J.C., Iliev.I., Johnson.R., Jones.C., Kann.L., Karatas.A.,
Klein.J., Lacroque.K., Lamazares.R., Landers.T., Lehoczy.J.,
Levine.R., Lieu.C., Liu.G., Locke.K., Macdonald.P., Marquis.N.,
McCarthy.M., McEwan.P., McGurk.A., McKernan.K., McPheeters.R.,

Meldrim.J., Meneus.L., Mihova.T., Miranda.C., Mlenga.V., Morrow.J.,
Murphy.T., Naylor.J., Norman.C.H., O'Connor.T., O'Donnell.P.,
O'Neil.D., Oliver.T.M., Oliver.J., Peterson.K., Pierre.N.,
Pisanu.C., Pollara.V., Raymond.C., Riley.R., Rogov.P., Rothman.D.,
Roy.A., Santos.R., Schauer.S., Severy.P., Spencer.B.,
Stange-Thomann.N., Stojanovic.N., Subramanian.A., Talamas.J.,
Tessaye.S., Theodore.J., Tirrell.A., Travers.M., Trigilio.J.,
Vassiliev.H., Viel.R., Vo.A., Wilson.B., Wu.X., Wyman.D., Ye.W.J.,
Young.G., Zainoun.J., Zimmer.A. and Zody.M.
Direct Submission
Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 8, 2001 this sequence version replaced gi:7677926.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9099
Center clone name: 359_A_12
----- Summary Statistics
Sequencing vector: M13; M77815; 35% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 191730 bases at least Q40
Consensus quality: 193959 bases at least Q30
Consensus quality: 194605 bases at least Q30
Insert size: 172000; agarose-fp
Insert size: 194961; sum-of-contigs
Quality coverage: 10.5 in Q20 bases; agarose-fp
Quality coverage: 9.3 in Q20.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 36680: contig of 36680 bp in length
* 36681 36780: gap of 100 bp
* 36781 38091: contig of 1311 bp in length
* 38092 38191: gap of 100 bp
* 38192 39416: contig of 1225 bp in length
* 39417 39516: gap of 100 bp
* 39517 40566: contig of 1050 bp in length
* 40567 40666: gap of 100 bp
* 40667 42013: contig of 1347 bp in length
* 42014 42113: gap of 100 bp
* 42114 43651: contig of 1538 bp in length
* 43652 43751: gap of 100 bp
* 43752 44775: contig of 1024 bp in length
* 44776 44875: gap of 100 bp
* 44876 195661: contig of 150786 bp in length.
Location/Qualifiers
1..195661
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosomes="8"
/map="8"
/clone="RP11-359A12"
/clone_lib="RPCI-11 Human Male BAC"
1..36680
/note="assembly_fragment
clone_end:SP6
vector_side:left"
36781..38091
/note="assembly_fragment"
FEATURES
Source
misc_feature
misc_feature


```
misc_feature 38192..39416
              /note="assembly_fragment"
misc_feature 39517..40566
              /note="assembly_fragment"
misc_feature 40667..42013
              /note="assembly_fragment"
misc_feature 42114..43651
              /note="assembly_fragment"
misc_feature 43752..44775
              /note="assembly_fragment"
misc_feature 44876..45661
              /note="assembly_fragment"
              clone_end:r7
              vector_side:right"
BASE COUNT 59915 a 40215 c 39680 g 55151 t 700 others
ORIGIN
-----
Query Match 1.3%; Score 73; DB 2; Length 195661;
Best Local Similarity 100.0%; Pred. No. 5.3e-28;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5439 GCTAATGTTTCTATTATTTAGTAGAGATGGGGTTTCACCATGTTGCCAGGCTGCTCAAA 5498
      |||||||
Db 51109 GCTAATGTTTCTATTATTTAGTAGAGATGGGGTTTCACCATGTTGCCAGGCTGCTCAAA 51168

Qy 5499 ACTCCTGACCTCA 5511
      |||||||
Db 51169 ACTCCTGACCTCA 51181

RESULT 15
LOCUS AF235101 214696 bp DNA linear HTG 26-JUL-2002
DEFINITION Homo sapiens chromosome 8 clone CTD-2571K21 map 8q, WORKING DRAFT
SEQUENCE, 11 unordered pieces.
ACCESSION AF235101
VERSION AF235101.4 GI:16356856
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_CANCELLED.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          Wen,G., Baumgart,C., Blechschmidt,K., Dette,M., Jahn,N.,
          Lehmann,R., Menzel,U., Polley,A., Reichwald,K., Schilhabel,M.B.,
          Schudy,A., Siddiqui,R., Taudien,S., Rosenthal,A. and Platzer,M.
          Chromosome 8 genomic sequence
          Unpublished
          2 (bases 1 to 214696)
          Blechschmidt,K., Wen,G., Schilhabel,M., Baumgart,C., Menzel,U.,
          Dette,M., Jahn,N. and Rosenthal,A.
          Direct Submission
          Submitted (16-FEB-2000) Genome Analysis, Institute of Molecular
          Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
          On Oct 24, 2001 this sequence version replaced gi:14329032.
          -----
          Center: Genome Center
          Center: Institute of Molecular Biotechnology
          Center code: IMB
          Web site: http://genome.imb-jena.de/
          Contact: gscj-submit@genome.imb-jena.de
          -----
          Project Information
          Center project name: H360
          Center clone name: CTD-2571K21
          -----
          Summary Statistics
          Sequencing vector: M13; 100% of reads
          Chemistry: Dye-terminator Big Dye; 100% of reads
          Assembly program: Phrap; version 0.990329
          Consensus quality: 200056 bases at least Q40
          Consensus quality: 206842 bases at least Q30
          Consensus quality: 210985 bases at least Q20
          Quality coverage: 4.93 x in Q20 bases; sum-of-contigs
          -----
          Sequence Quality Assessment:
```

```
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality 10.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 14268: contig of 14268 bp in length
* 14269 14368: gap of unknown length
* 14369 16329: contig of 1961 bp in length
* 16330 16429: gap of unknown length
* 16430 21020: contig of 4591 bp in length
* 21021 21120: gap of unknown length
* 21121 31366: contig of 10246 bp in length
* 31367 31467: gap of unknown length
* 31467 38824: contig of 7358 bp in length
* 38825 38925: gap of unknown length
* 38925 48608: contig of 9684 bp in length
* 48609 48708: gap of unknown length
* 48709 67136: contig of 18428 bp in length
* 67137 67237: gap of unknown length
* 67237 138505: contig of 71269 bp in length
* 138506 138605: gap of unknown length
* 138606 168896: contig of 28291 bp in length
* 168897 168996: gap of unknown length
* 168997 197031: contig of 30035 bp in length
* 197032 197131: gap of unknown length
* 197132 214696: contig of 17565 bp in length.
          Location/Qualifiers
            1..214696
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="8"
              /map="8q"
              /clone="CTD-2571K21"
BASE COUNT 63313 a 42987 c 42987 g 64499 t 1000 others
ORIGIN
-----
Query Match 1.3%; Score 73; DB 2; Length 214696;
Best Local Similarity 100.0%; Pred. No. 5.3e-28;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5439 GCTAATGTTTCTATTATTTAGTAGAGATGGGGTTTCACCATGTTGCCAGGCTGCTCAAA 5498
      |||||||
Db 172295 GCTAATGTTTCTATTATTTAGTAGAGATGGGGTTTCACCATGTTGCCAGGCTGCTCAAA 172354

Qy 5499 ACTCCTGACCTCA 5511
      |||||||
Db 172355 ACTCCTGACCTCA 172367

Search completed: June 19, 2003, 17:42:00
Job time : 13765 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 00:37:34 ; Search time 1071 Seconds
(without alignments)
11594.322 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 5514
Sequence: 1 acagacgaatacataggtcca.....tcaaacctcctgacctcagag 5514

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID52/qcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID52/qcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID52/qcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID52/qcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID52/qcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID52/qcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID52/qcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID52/qcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID52/qcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID52/qcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID52/qcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID52/qcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID52/qcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID52/qcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID52/qcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID52/qcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID52/qcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID52/qcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID52/qcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID52/qcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID52/qcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID52/qcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID52/qcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID52/qcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5514	100.0	5514	21 AAC55313	Human activation-i
2	4771	86.5	11204	21 AAC55339	Human activation-i
c 3	118	2.1	772	23 AAS81193	DNA encoding novel
4	87	1.6	87	21 AAC55315	Human activation-i
5	87	1.6	2818	21 AAC55312	Human activation-i
c 6	71	1.3	12718	24 ABK52229	CDNA encoding huma
c 7	69	1.3	539	23 ABV57689	Human prostate exp
c 8	67	1.2	84	22 AAL07192	Human reproductive
9	67	1.2	95	22 ABA18401	Human nervous syst

c 10	67	1.2	121	22 ABA14482	Human nervous syst
c 11	67	1.2	132	22 AAK80307	Human immune/haema
c 12	67	1.2	141	22 AAK66840	Human immune/haema
c 13	67	1.2	141	22 AAK68098	Human immune/haema
c 14	67	1.2	141	22 AAK68878	Human immune/haema
c 15	67	1.2	141	22 AAK70059	Human immune/haema
c 16	67	1.2	141	22 AAK72514	Human immune/haema
c 17	67	1.2	141	22 AAK72808	Human immune/haema
c 18	67	1.2	141	22 AAK72892	Human immune/haema
c 19	67	1.2	141	22 AAK80869	Human immune/haema
c 20	67	1.2	141	22 AAK83523	Human immune/haema
c 21	67	1.2	141	22 AAK85680	Human immune/haema
c 22	67	1.2	161	22 AAK66827	Human immune/haema
c 23	67	1.2	481	22 ABA18403	Human nervous syst
24	67	1.2	566	22 AAH10398	Human cDNA clone (
c 25	67	1.2	606	24 ABQ57438	Human colon cancer
26	67	1.2	1160	22 AAL02996	Human reproductive
27	67	1.2	1160	23 ABL97343	Human testicular a
c 28	67	1.2	1607	22 AAH17521	Human cDNA sequenc
c 29	67	1.2	3015	22 AAK85238	Human immune/haema
30	67	1.2	3780	22 AAK72236	Human immune/haema
31	67	1.2	3780	22 AAK79330	Human immune/haema
32	67	1.2	4736	21 AAC69133	Human ABC1 gene ex
c 33	67	1.2	5930	22 AAL36019	Human musculoskele
34	67	1.2	7232	22 AAK72237	Human immune/haema
c 35	67	1.2	7232	22 AAK79332	Human immune/haema
c 36	67	1.2	9139	22 AAK80305	Human immune/haema
c 37	67	1.2	9139	22 AAK80306	Human immune/haema
c 38	67	1.2	9723	22 AAK65237	Human immune/haema
c 39	67	1.2	11426	22 AAK84781	Human immune/haema
c 40	67	1.2	11754	22 AAS04033	Human ABC1 gene, p
c 41	67	1.2	16086	22 AAL36020	Human musculoskele
c 42	67	1.2	16086	22 AAK69721	Human immune/haema
c 43	67	1.2	16869	21 ABN97976	Human retroviral s
c 44	67	1.2	20303	18 AAT71699	Human deoxycytidyl
c 45	67	1.2	23580	22 AAK66230	Human immune/haema

ALIGNMENTS

RESULT 1

AAC55313
ID AAC55313 standard; DNA; 5514 BP.

XX AAC55313;

AC AAC55313;

XX 05-FEB-2001 (first entry)

DT Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:9.

DE Activation-induced cytidine deaminase; AID; cytidine deaminase;
KW immune related disease; allergy; allergic disease; anti-allergic;
KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;
KW gene therapy; B cell associated immune system disorder; food allergy;
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;
KW drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;
KW ataxia telangiectasia; common variable immunodeficiency disorder;
KW major histocompatibility class II deficiency disease;
KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.

OS Homo sapiens.

XX WO200058480-A1.

PD 05-OCT-2000.

PF 28-MAR-2000; 2000WO-JP01918.

XX 29-MAR-1999; 99JP-0087192.

PR 24-JUN-1999; 99JP-0178999.

PR 27-DEC-1999; 99JP-0371382.

1621 AAATGCCAAGCATTCATAATTAATGCTTGAAAGTCACTATGATGCTGCCATTATAAGG 1680
1621 AAATGCCAAGCATTCATAATTAATGCTTGAAAGTCACTATGATGCTCCATTATAAGG 1680
1681 AGACAAATTCATTCAAGCAAGTATATTAATGTTAAAGGCCCAATGTTAGGCAGTTAATG 1740
1681 AGACAAATTCATTCAAGCAAGTATATTAATGTTAAAGGCCCAATGTTAGGCAGTTAATG 1740
1741 GCACCTTTTACTATTAACTTAATCTTCCATTTGTTCCAGAGCTAGCTTAACCTTACCTTTAG 1800
1741 GCACCTTTTACTATTAACTTAATCTTCCATTTGTTCCAGAGCTAGCTTAACCTTACCTTTAG 1800
1801 GTGTGAATTTGGTTAAGTTCCTCATPAATATGCTTTATGTGCAGTGTGTTTATAGGTTATTGT 1860
1801 GTGTGAATTTGGTTAAGTTCCTCATPAATATGCTTTATGTGCAGTGTGTTTATAGGTTATTGT 1860
1861 CATAGAACTTATCTTATTCCTACATTTTATGATTACTATGGATGGATGATGAGAATAACACCTA 1920
1861 CATAGAACTTATCTTATTCCTACATTTTATGATTACTATGGATGGATGATGAGAATAACACCTA 1920
1921 ATCCTTATACTTTTACTCCTCAATTTTACTCCTTTTATAAAGAATCTTACATTTACAGAATAAAGA 1980
1921 ATCCTTATACTTTTACTCCTCAATTTTACTCCTTTTATAAAGAATCTTACATTTACAGAATAAAGA 1980
1981 TTTTTTAAAAATATATTTTTTTGTAGACAGGGTCTTTAGCCAGCCGAGGCTGTCTCT 2040
1981 TTTTTTAAAAATATATTTTTTTGTAGACAGGGTCTTTAGCCAGCCGAGGCTGTCTCT 2040
2041 AAGTCTTGCCCAAGCGATCTCTGCTGCTGGCCCTTAAAGTGTGCGAATATATAGACAT 2100
2041 AAGTCTTGCCCAAGCGATCTCTGCTGCTGGCCCTTAAAGTGTGCGAATATATAGACAT 2100
2101 GAGCCATCACATCCAAATACAGAATAAAGATTTTTTAAATGAGGAGTTTAAATGCTTCTTCAG 2160
2101 GAGCCATCACATCCAAATACAGAATAAAGATTTTTTAAATGAGGAGTTTAAATGCTTCTTCAG 2160
2161 AAAATTTTCTTGAGTGCAGACAAATGTCAAATGTCTCTCAGTTTACACTGAGATTTTGAA 2220
2161 AAAATTTTCTTGAGTGCAGACAAATGTCAAATGTCTCTCAGTTTACACTGAGATTTTGAA 2220
2221 RACAAGTCTGAGCTATAGGTCCTTGTGGAAGGTCATTTGGAATACTTGTTCAAAGTAAA 2280
2221 RACAAGTCTGAGCTATAGGTCCTTGTGGAAGGTCATTTGGAATACTTGTTCAAAGTAAA 2280
2281 ATGGAAGCAAAAGGTAATCAGCAGTTTGAATTCAGAGAAAGACAGAAAAAGGAGAAAAG 2340
2281 ATGGAAGCAAAAGGTAATCAGCAGTTTGAATTCAGAGAAAGACAGAAAAAGGAGAAAAG 2340
2341 ATGAAATTCACAGGACAGAAAGGGAATATATTTATCATTAAGGAGGACAGTATCTGTAGA 2400
2341 ATGAAATTCACAGGACAGAAAGGGAATATATTTATCATTAAGGAGGACAGTATCTGTAGA 2400
2401 GCTCATTTAGTATGTCGCAAAATGACTTGTGTCAGGATTTATTTTAAACCCGCTTGTTCGTGT 2460
2401 GCTCATTTAGTATGTCGCAAAATGACTTGTGTCAGGATTTATTTTAAACCCGCTTGTTCGTGT 2460
2461 TTGCAGGCTGGGATGACAGTGTGCTCCTCAGGAGCAGACAGCTGTCAGAGCAGC 2520
2461 TTGCAGGCTGGGATGACAGTGTGCTCCTCAGGAGCAGACAGCTGTCAGAGCAGC 2520
2521 TGTACGCTGGAAGCCTGAAACACTCCCTCGGTAAAGTCCCTTCTTACTCAGGACAGAAAT 2580
2521 TGTACGCTGGAAGCCTGAAACACTCCCTCGGTAAAGTCCCTTCTTACTCAGGACAGAAAT 2580
2581 GACGAGAACAGGAGCTGGAACAGGCCCTTAAACAGAGAGGGGAGTAAATGGATCAACA 2640
2581 GACGAGAACAGGAGCTGGAACAGGCCCTTAAACAGAGAGGGGAGTAAATGGATCAACA 2640
2641 AAGTTTAACTAGCAGGTTCAGGATCAGCAATTTCAATTTTCACTCTGTACTGTTAATCATCTGTACA 2700
2641 AAGTTTAACTAGCAGGTTCAGGATCAGCAATTTCAATTTTCACTCTGTACTGTTAATCATCTGTACA 2700
2701 GAAACAGTGTAGGCTTATTGTTTTCATGTAGAGTAGGACCCCAAAAAATCCACCCAAAGT 2760

2701 GAAACAGTGTAGGCTTATTGTTATTTTCATGTAGAGTAGGACCCCAAAAAATCCACCCAAAGT 2760
2761 CCTTTATCTATGCCACATCCCTTCTTATCTATACTTCCAGGACACTTTTCTTCTCTATGA 2820
2761 CCTTTATCTATGCCACATCCCTTCTTATCTATACTTCCAGGACACTTTTCTTCTCTATGA 2820
2821 TAAAGCTCTCTCTCTCCAC 2880
2821 TAAAGCTCTCTCTCTCCAC 2880
2881 CACAAAC 2940
2881 CACAAAC 2940
2941 CTCATCTACACAGCCCGAGGAGGTAAGTTAATATAGAGGGATTTATTTGGTAAAGATGA 3000
2941 CTCATCTACACAGCCCGAGGAGGTAAGTTAATATAGAGGGATTTATTTGGTAAAGATGA 3000
3001 TGCCTTAACTCTGTTTAAACACTGGGCTCAAGAGAGAAATTTCTTTCTCTGCTACTATTA 3060
3001 TGCCTTAACTCTGTTTAAACACTGGGCTCAAGAGAGAAATTTCTTTCTCTGCTACTATTA 3060
3061 AGCACCATTATGCTTGGAGCTTATATACAAAGGGTATTTATATGCTAATATAGTAAT 3120
3061 AGCACCATTATGCTTGGAGCTTATATACAAAGGGTATTTATATGCTAATATAGTAAT 3120
3121 AGTAATGTGGTGGTGTACTATGGTAAATTTACCATAAATAATTAATATCTTTTAAAAATAAG 3180
3121 AGTAATGTGGTGGTGTACTATGGTAAATTTACCATAAATAATTAATATCTTTTAAAAATAAG 3180
3181 CTAATTAATTTTGGATCTTTTCTAGTATTTTCTATTTTATGTTTATGTTTGTGATTTT 3240
3181 CTAATTAATTTTGGATCTTTTCTAGTATTTTCTATTTTATGTTTATGTTTGTGATTTT 3240
3241 AAAAGACAATCTCACCTCTTACCAGGCTGGAGTGCAGTGGTGCAGTGCAGTGCAGTGCAGT 3300
3241 AAAAGACAATCTCACCTCTTACCAGGCTGGAGTGCAGTGGTGCAGTGCAGTGCAGTGCAGT 3300
3301 CAGCTTGAACCTCTGGGCTCAAGCAATCTCCTGCTGGCTTGGCTTGGCTTGGCTTGGCTTGG 3360
3301 CAGCTTGAACCTCTGGGCTCAAGCAATCTCCTGCTGGCTTGGCTTGGCTTGGCTTGGCTTGG 3360
3361 ACAGTCATGAGCAGCTGCTGCTAGGATCCATTTAGATTAATAATATGATTTTAA 3420
3361 ACAGTCATGAGCAGCTGCTGCTAGGATCCATTTAGATTAATAATATGATTTTAA 3420
3421 TTTTAAAAATAATGCTTAAATTTTACCTTTATGTTAATGTGTATGCTGCTTAAATCTAG 3480
3421 TTTTAAAAATAATGCTTAAATTTTACCTTTATGTTAATGTGTATGCTGCTTAAATCTAG 3480
3481 TTTGCTGCTTAAAGTTTAAAGTGCCTTCCAAATAGCTTCAATGCTGCTGCTGCTGCTGCTGCT 3540
3481 TTTGCTGCTTAAAGTTTAAAGTGCCTTCCAAATAGCTTCAATGCTGCTGCTGCTGCTGCTGCT 3540
3541 TAAAGTGAACAGACAGCAGGCTGGTGGCTCAGCCTGTAATCCAGCAGCTCTGGGAG 3600
3541 TAAAGTGAACAGACAGCAGGCTGGTGGCTCAGCCTGTAATCCAGCAGCTCTGGGAG 3600
3601 GCTGAGGTGGTGGATTCGCTTGGAGCCCTGGAGTTCAAGACAGCCTGAGCAACATGGCAA 3660
3601 GCTGAGGTGGTGGATTCGCTTGGAGCCCTGGAGTTCAAGACAGCCTGAGCAACATGGCAA 3660
3661 AACCTGTTCTATATAAATAATTTAGCCGGCATGGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3720
3661 AACCTGTTCTATATAAATAATTTAGCCGGCATGGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3720
3721 ACTAGGGGCTGAGGACAGAGATCTTTGGAGCCAGAGGTCAGAGCTGCACTGAGCAG 3780
3721 ACTAGGGGCTGAGGACAGAGATCTTTGGAGCCAGAGGTCAGAGCTGCACTGAGCAG 3780
3781 TGCCTTGCCTTCTGCACTCCAGCCTGGGTGACAGGACCAAGACCTTTGCCCTCAAAAAATAA 3840
3781 TGCCTTGCCTTCTGCACTCCAGCCTGGGTGACAGGACCAAGACCTTTGCCCTCAAAAAATAA 3840

Db	3781	TGCTTGGCCACCTGCACCTCAGCCTGGGTGACAGGACCAGACCTTGCCCTCAAAAAATAA	3840	QY	4921	CAAGACACTGCTAGGAGCCAGAAACAAAGAGGAGGAGAAATCAGTCATTATGTGGGA	4980
QY	3841	GAAGAAAAATTAATAATTAATGAACAACTACAAAGAGCTGTGTCCCTAGATGAGCTAC	3900	Db	4921	CAAGACACTGCTAGGAGCCAGAAACAAAGAGGAGGAGAAATCAGTCATTATGTGGGA	4980
Db	3841	GAAGAAAAATTAATAATTAATGAACAACTACAAAGAGCTGTGTCCCTAGATGAGCTAC	3900	QY	4981	ACAACATAGCAAGATATTTAGATCATTTTGACTAGTTTAAAAAAACGACGACAGTACAAAAT	5040
QY	3901	TTAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGCTGTGCACCTGCACCTAC	3960	Db	4981	ACAACATAGCAAGATATTTAGATCATTTTGACTAGTTTAAAAAAACGACGACAGTACAAAAT	5040
Db	3901	TTAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGCTGTGCACCTGCACCTAC	3960	QY	5041	CACACATGCATCAGTATATTAATCCAAATCATGTAAATATGTGCCTGTAGAAAAGACTAGAGG	5100
QY	3961	ATTATTAAATATCAATCTCAATGTATATCCACACAAAGACATGGTACGTGAATGTTCAT	4020	Db	5041	CACACATGCATCAGTATATTAATCCAAATCATGTAAATATGTGCCTGTAGAAAAGACTAGAGG	5100
Db	3961	ATTATTAAATATCAATCTCAATGTATATCCACACAAAGACATGGTACGTGAATGTTCAT	4020	QY	5101	AATAAACACAGAATCTTAAACAGTCATTGTCAATTAGACACTAAGTCTAAATTATTATTATT	5160
QY	4021	AGTACCTTTATTCCAAAAACCCCAAGTAGAGACTATCCAAATATCCATCAACAAGTGAA	4080	Db	5101	AATAAACACAGAATCTTAAACAGTCATTGTCAATTAGACACTAAGTCTAAATTATTATTATT	5160
Db	4021	AGTACCTTTATTCCAAAAACCCCAAGTAGAGACTATCCAAATATCCATCAACAAGTGAA	4080	QY	5161	AGACACTATGATATTTGAGATTTTAAAAATCTTTTAAATATTTTAAAAATTTAGAGCTCTTCT	5220
QY	4081	CAATTAACAAATGTCTATATCCATGCAATGGAATACACCCCTGCAGTACAAAGGAAG	4140	Db	5161	AGACACTATGATATTTGAGATTTTAAAAATCTTTTAAATATTTTAAAAATTTAGAGCTCTTCT	5220
Db	4081	CAATTAACAAATGTCTATATCCATGCAATGGAATACACCCCTGCAGTACAAAGGAAG	4140	QY	5221	ATTTTCCATAGTATTTCAAGTTTGAACAATGATCAAGTATTTACTCTTTCTTTTCTTTT	5280
QY	4141	AACTACTTTGGGATGAATCCCAAAAGTCATGACGCTAAATGAAGAGTCAGACATCAAGG	4200	Db	5221	ATTTTCCATAGTATTTCAAGTTTGAACAATGATCAAGTATTTACTCTTTCTTTTCTTTT	5280
Db	4141	AACTACTTTGGGATGAATCCCAAAAGTCATGACGCTAAATGAAGAGTCAGACATCAAGG	4200	QY	5281	TTTTTTTTTTTTGAGATGGAGTTTGGTCTTGTGCCATGCTGGAGTGAATGGCATG	5340
QY	4201	AGGAGATAATGTATGCCATACGAAATTTCTAGAAAATGAAAGTAACTTATAGTTACAGAAA	4260	Db	5281	TTTTTTTTTTTTGAGATGGAGTTTGGTCTTGTGCCATGCTGGAGTGAATGGCATG	5340
Db	4201	AGGAGATAATGTATGCCATACGAAATTTCTAGAAAATGAAAGTAACTTATAGTTACAGAAA	4260	QY	5341	AYCATAGCTCAGTCGCAACCTCCACCTCTCGGGTTCAAGCAAAAGCTGTGCGCTCAGCCTCC	5400
QY	4261	GCAATCAGGGCAGGATAGAGCTCACACCTGTAATCCAGACATTTTGAGAGGCCACGT	4320	Db	5341	AYCATAGCTCAGTCGCAACCTCCACCTCTCGGGTTCAAGCAAAAGCTGTGCGCTCAGCCTCC	5400
Db	4261	GCAATCAGGGCAGGATAGAGCTCACACCTGTAATCCAGACATTTTGAGAGGCCACGT	4320	QY	5401	CGGGTAGATGGGATTACAGCGCCACCACCACTCGGCTAAATGTTGTATTTTGTATAGTA	5460
QY	4321	GGGAAGATGCTAGAACTCAGGAGTTCAAGACCAGCTTGGGCAACACAGTGAACCTCCAT	4380	Db	5401	CGGGTAGATGGGATTACAGCGCCACCACCACTCGGCTAAATGTTGTATTTTGTATAGTA	5460
Db	4321	GGGAAGATGCTAGAACTCAGGAGTTCAAGACCAGCTTGGGCAACACAGTGAACCTCCAT	4380	QY	5461	GAGATGGGTTTCCACCATTGTTGCCAGGCTGGTCTCAAACTCCTGACCTCAGAG	5514
QY	4381	TTCCCAAAAAATGGAAAAAAGAAAGCAAAATCAGTGGTGTGCTGTGGGAGGGGAAG	4440	Db	5461	GAGATGGGTTTCCACCATTGTTGCCAGGCTGGTCTCAAACTCCTGACCTCAGAG	5514
Db	4381	TTCCCAAAAAATGGAAAAAAGAAAGCAAAATCAGTGGTGTGCTGTGGGAGGGGAAG	4440	RESULT 2			
QY	4441	GACTGCAAGAGGGAAGAGCTCTGTGGGTGAGGGTGGTATTCAGGTTCTGTATCCT	4500	AAC55339			
Db	4441	GACTGCAAGAGGGAAGAGCTCTGTGGGTGAGGGTGGTATTCAGGTTCTGTATCCT	4500	ID	AAC55339	standard; DNA; 11204 BP.	
QY	4501	GACTGTGTAGCAGTTTGGGGTGTTTACATCCAAAAATATTCGTAGAATTTATGCATCTTA	4560	XX	AC	AAC55339;	
Db	4501	GACTGTGTAGCAGTTTGGGGTGTTTACATCCAAAAATATTCGTAGAATTTATGCATCTTA	4560	XX	AC	AAC55339;	
QY	4561	ATGGGTGAGTTTACTGTATGTAATATACCTCAATGTAAGAAAAATAATGTGTAAG	4620	XX	DT	05-FEB-2001 (first entry)	
Db	4561	ATGGGTGAGTTTACTGTATGTAATATACCTCAATGTAAGAAAAATAATGTGTAAG	4620	XX	DE	Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:35.	
QY	4621	AAAAGTTTCAATCTCTCCAGCAACCTTATCAAAATCTTGAGCCCTTTACTTCGCA	4680	XX	DE	Activation-induced cytidine deaminase; AID; cytidine deaminase;	
Db	4621	AAAAGTTTCAATCTCTCTCCAGCAACCTTATCAAAATCTTGAGCCCTTTACTTCGCA	4680	XX	DE	Immune related disease; allergy; allergic disease; antiallergic;	
QY	4681	AATTTCTCTCACTTCTGCCCTTACCATTTAGTGACAGCACTAGCTCCACAAATTTGGATA	4740	XX	DE	antianemic; antidiabetic; ophthalmological; anti-HIV; dermatological;	
Db	4681	AATTTCTCTCACTTCTGCCCTTACCATTTAGTGACAGCACTAGCTCCACAAATTTGGATA	4740	XX	DE	gene therapy; B cell associated immune system disorder; food allergy;	
QY	4741	AATGCATTTCTGGAAAAAGACTAGGACAAAAATCCAGGCATCACTTGTGCTTTTATATCAA	4800	XX	DE	immunodeficiency disease; immunoglobulin A deficiency disease; asthma;	
Db	4741	AATGCATTTCTGGAAAAAGACTAGGACAAAAATCCAGGCATCACTTGTGCTTTTATATCAA	4800	XX	DE	IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;	
QY	4801	CCAGCTGTACAGCTGTGCTCTCTCCAGCTGCAATGGGACTCTTGATTTCTTTAA	4860	XX	DE	drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;	
Db	4801	CCAGCTGTACAGCTGTGCTCTCTCCAGCTGCAATGGGACTCTTGATTTCTTTAA	4860	XX	DE	ataxia telangiectasia; common variable immunodeficiency disorder;	
QY	4861	GGAAACTTGGGTACCAGATATTTCCAAAAATGCTATTCAAATTAGTCTTATGATATG	4920	XX	OS	major histocompatibility class II deficiency disease;	
Db	4861	GGAAACTTGGGTACCAGATATTTCCACAAATGCTATTCAAATTAGTCTTATGATATG	4920	XX	OS	auto immunodeficiency syndrome; IgG subclass selection disorder; ds.	
				XX	OS	Homo sapiens.	
				XX	PN	WO200058480-A1.	
				XX	PN		
				XX	PD	05-OCT-2000.	
				XX	PF	28-MAR-2000; 2000WO-JP01918.	
				XX	PR	29-MAR-1999; 99JP-0087192.	
				PR	PR	24-JUN-1999; 99JP-0178999.	

PR 27-DEC-1999; 99JP-0371382.
XX (NIBS) JAPAN TOBACCO INC.
PA (HONJ/) HONJO T.
XX
PI Honjo T, Muramatsu M;
XX WPI; 2000-611715/58.
DR
XX Nucleic acid encoding activation induced cytidine deaminase, useful as
PT a target for drug development for immune-related diseases including
PT allergies -
XX
PS Claim 17; Page 163-170; 174pp; Japanese.
XX
CC The present invention describes an activation-induced cytidine deaminase
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and
CC has cytidine activity similar to APOBEC-1. AID has antiallergic,
CC antianalemic, antiasthmatic, ophthalmological, anti-HIV and
CC dermatological activities, and can be used in gene therapy. AID
CC polynucleotides are useful in methods for identifying drugs for the
CC treatment of B cell associated immune system disorders, immunodeficiency
CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic
CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen
CC disease, DiGeorge disease, ataxia telangiectasia, common variable
CC immunodeficiency disorder, MHC (major histocompatibility class) class
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
CC IgE disorder, and IgG subclass selection disorder. The DNA sequences
CC encoding AID may be used for gene therapy and the antibodies to the AID
CC protein may be used for diagnosis and treatment of these disorders. The
CC present sequence represents a genomic DNA sequence of human AID.
XX
SQ Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 other;

Query Match 86.5%; Score 4771; DB 21; Length 11204;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 4921; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 591 AGGTTTCAGAGAGCTGGGAATATGGGGGAATAGAGGCTATCTGAGGCTCTTCAACAC 650
DB 1 AGGTTTCAGAGAGCTGGGAATATGGGGGAATAGAGGCTATCTGAGGCTCTTCAACAC 60
QY 651 AATAACCAAGAAGCTATTAAATGCTCTTTAAAGGTATTATACATAAATATTACTATTCTC 710
DB 61 AATAACCAAGAAGCTATTAAATGCTCTTTAAAGGTATTATACATAAATATTACTATTCTC 120
QY 711 ATTGTGCTTTATTTTGTGTTATCATGATTATATTAATGAAGTGCTACTGTTACTGCCPCC 770
DB 121 ATTGTGCTTTATTTTGTGTTATCATGATTATATTAATGAAGTGCTACTGTTACTGCCPCC 180
QY 771 TGATCTTTGCTAGCTATGAGCATGGAGCTGGGCTTTTATAGCAGCAGCCCCAAAGGAACC 830
DB 181 TGATCTTTGCTAGCTATGAGCATGGAGCTGGGCTTTTATAGCAGCAGCCCCAAAGGAACC 240
QY 831 TAAACATTAAAGCAGAGCTGCCCTCAATGGTTTAAACCTGTGTGACTCTGCCCTATGACAGC 890
DB 241 TAAACATTAAAGCAGAGCTGCCCTCAATGGTTTAAACCTGTGTGACTCTGCCCTATGACAGC 300
QY 891 CCCACCCACCCTTCTACTGGATCCAAATCAGGAGCAAGGCCGTTGGGGTACCTGTGGTG 950
DB 301 CCCACCCACCCTTCTACTGGATCCAAATCAGGAGCAAGGCCGTTGGGGTACCTGTGGTG 360
QY 951 GGGTGAGCTCTCAGGGGAGAGGCCCAAAAGGGCAAGCTCAAAATTTGAATGTGAAGGCC 1010
DB 361 GGGTGAGCTCTCAGGGGAGAGGCCCAAAAGGGCAAGCTCAAAATTTGAATGTGAAGGCC 420
QY 1011 AATGCACTGTGAGACTGAGACAGAACCATCATTAATTGAAGTGAGATTTTCTGGCCT 1070
DB 421 AATGCACTGTGAGACTGAGACAGAACCATCATTAATTGAAGTGAGATTTTCTGGCCT 480
QY 1071 GAGACTTGCAGGGAGGCAAGAGACACTCTCGACACCACCTATGGACAGGTAAGAGGCAG 1130
|||||

DB 481 GAGACTTGCAGGGAGGCAAGAGACACACTCTGGACACCACACTATGGACAGGTAAGAGGCAG 540
QY 1131 TCTTCTCGTGGGTGATTCGACTGGCCTTCCCTCTCAGAGCAAAATCTGAGTAATGACACTGG 1190
DB 541 TCTTCTCGTGGGTGATTCGACTGGCCTTCCCTCTCAGAGCAAAATCTGAGTAATGAGACTGG 600
QY 1191 TAGCTATCCCTTTCTCTCATGTACTGACTGACTGATAAGATCACTTGAATCAATATGCA 1250
DB 601 TAGCTATCCCTTTCTCTCATGTACTGACTGATAAGATCACTTGAATCAATATGCA 660
QY 1251 TATATATTTTGTGATCTGCTCTCTTTCTTATTCAGATCTTATPACGCTGTCAGCCAA 1310
DB 661 TATATATTTTGTGATCTGCTCTCTTTCTTATTCAGATCTTATPACGCTGTCAGCCAA 720
QY 1311 TTCCTTCGTTCAGACTTCTCTGATTTCCCTCTTTTTCATGTGGCAAAAGACTAGTG 1370
DB 721 TTCCTTCGTTCAGACTTCTCTGATTTCCCTCTTTTCATGTGGCAAAAGACTAGTG 780
QY 1371 CGTACAATGTACTGATTCGCTCTCGAGATTTGTACCATGTTTGAAACTAAATTTATGTTAT 1430
DB 781 CGTACAATGTACTGATTCGCTCTCGAGATTTGTACCATGTTTGAAACTAAATTTATGTTAT 840
QY 1431 AATATTAACATAGCAAAATCTTTAGAGACTCAAAATCATGAAAAGGTAATAGCAGTACTGTA 1490
DB 841 AATATTAACATAGCAAAATCTTTAGAGACTCAAAATCATGAAAAGGTAATAGCAGTACTGTA 900
QY 1491 CTAATAACGGTAGTCTAAATTTTCGTAATAATTTTGTAAATTTCAACAGTAACAACT 1550
DB 901 CTAATAACGGTAGTCTAAATTTTCGTAATAATTTTGTAAATTTCAACAGTAACAACT 960
QY 1551 TGAAGACACACTTCTCTAGGGAGGGTACTGAAATAATTTAGCTTATAGTAAGAAATTT 1610
DB 961 TGAAGACACACTTCTCTAGGGAGGGTACTGAAATAATTTAGCTTATAGTAAGAAATTT 1020
QY 1611 GTAATTTTAGAAATGCCAAGCATTTCTAAATTAATTTGCTTGAAGTCACTATGATGTGTC 1670
DB 1021 GTAATTTTAGAAATGCCAAGCATTTCTAAATTAATTTGCTTGAAGTCACTATGATGTGTC 1080
QY 1671 CATATATAGGAGACAAATTCATCAAGCAAGTATTATTAATGTTAAAGGCCCAATTTGTTAG 1730
DB 1081 CATATATAGGAGACAAATTCATCAAGCAAGTATTATTAATGTTAAAGGCCCAATTTGTTAG 1140
QY 1731 GCAGTTAATGGCAGCTTTTACTATTAACTAAATCTTTTCCATTTTGTTCAGACGTAGCTTAAT 1790
DB 1141 GCAGTTAATGGCAGCTTTTACTATTAACTAAATCTTTTCCATTTTGTTCAGACGTAGCTTAAT 1200
QY 1791 TACCTCTTAGGTGGAATTTGGTTAAGTCTCTATAATGCTTTTATGTCAGTCTTTTGTAT 1850
DB 1201 TACCTCTTAGGTGGAATTTGGTTAAGTCTCTATAATGCTTTTATGTCAGTCTTTTGTAT 1260
QY 1851 AGGTTATGTCATAGAACTTATCTATTCTTACATTTTATGATTAATGATGATGATGATGAGA 1910
DB 1261 AGGTTATGTCATAGAACTTATCTATTCTTACATTTTATGATTAATGATGATGATGAGA 1320
QY 1911 ATAACACCTAATCTCTTACTTACTCTCAATTTTAACTCTCTTTATATAAGAACTTACATTAC 1970
DB 1321 ATAACACCTAATCTCTTACTTACTCTCAATTTTAACTCTCTTTATATAAGAACTTACATTAC 1380
QY 1971 AGAATAAAGATTTTAAAAATATATTTTTTTTGTAGACAGAGGTCCTTAGCCCAAGCCGAG 2030
DB 1381 AGAATAAAGATTTTAAAAATATATTTTTTTTGTAGACAGAGGTCCTTAGCCCAAGCCGAG 1440
QY 2031 GCTGGTCTTAAGTCTTGGCCCAAGCATCTCTCTGCTGGGCTTCTTAAAGTGTCTGGA 2090
DB 1441 GCTGGTCTTAAGTCTTGGCCCAAGCATCTCTCTGCTGGGCTTCTTAAAGTGTCTGGA 1500
QY 2091 TTATAGACATGAGCCATCACATCAATATACAGATAAAGATTTTAAATGAGGATTTAA 2150
DB 1501 TTATAGACATGAGCCATCACATCAATATACAGATAAAGATTTTAAATGAGGATTTAA 1560
QY 2151 TGTCTTTCAGAAAAATTTTCTTGTAGGTGAGCAAAATGTCAAAATGTCCTCAGTCTTACACTG 2210
DB 1561 TGTCTTTCAGAAAAATTTTCTTGTAGGTGAGCAAAATGTCAAAATGTCCTCAGTCTTACACTG 1620

QY	2211	AGATTTTGAACAAGCTCTGAGCTATAGGTTCCTTGTGAAGGTCCTCATTTGGAATACTTGT	2270
Db	1621	AGATTTTGAACAAGCTCTGAGCTATAGGTTCCTTGTGAAGGTCCTCATTTGGAATACTTGT	1680
QY	2271	TCNAAGTAAATGGAAGCAAGGTAAAATACGACGTTGAAATTCAGAGAAAGACAGAAA	2330
Db	1681	TCNAAGTAAATGGAAGCAAGGTAAAATACGACGTTGAAATTCAGAGAAAGACAGAAA	1740
QY	2331	AGGAGAAAAGATGAATTCACACGAGCAGCAAGGGAATATATATCATTAAGGAGGACAG	2390
Db	1741	AGGAGAAAAGATGAATTCACACGAGCAGCAAGGGAATATATATCATTAAGGAGGACAG	1800
QY	2391	TATCTGTAGAGCTCATTAGTGTATGGCAAAATGACCTTGGTCAGGATTAATTTAAACCCGCT	2450
Db	1801	TATCTGTAGAGCTCATTAGTGTATGGCAAAATGACCTTGGTCAGGATTAATTTAAACCCGCT	1860
QY	2451	TGTTTCTGGTTGTCACGGCTGGGATGACGCTAGGTTCTCGCTCAGGAGCAGACGCTGT	2510
Db	1861	TGTTTCTGGTTGTCACGGCTGGGATGACGCTAGGTTCTCGCTCAGGAGCAGACGCTGT	1920
QY	2511	CCAGACAGCTGTCAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCCTTCCCTACTCA	2570
Db	1921	CCAGACAGCTGTCAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCCTTCCCTACTCA	1980
QY	2571	GGACAGAAATGACGAGAACAGGAGCTGGAACAGGCCCTTACCCAGAGAGGGAAGTAA	2630
Db	1981	GGACAGAAATGACGAGAACAGGAGCTGGAACAGGCCCTTACCCAGAGAGGGAAGTAA	2040
QY	2631	TGGATCAACAAAGTTAACTAGCAGGTCAGGATCAGCAATTCATTTCACTCTGACGTGGTA	2690
Db	2041	TGGATCAACAAAGTTAACTAGCAGGTCAGGATCAGCAATTCATTTCACTCTGACGTGGTA	2100
QY	2691	ACATGTGACAGAAACAGTGTAGGCTTATTTGTTATTTTCATGTAGTAGGACCCAAAAATC	2750
Db	2101	ACATGTGACAGAAACAGTGTAGGCTTATTTGTTATTTTCATGTAGTAGGACCCAAAAATC	2160
QY	2751	CACCCAAAGTCCTTTATCTATGCCACATCCTTCTTATCTATATCTTCCAGGACACTTTTTC	2810
Db	2161	CACCCAAAGTCCTTTATCTATGCCACATCCTTCTTATCTATATCTTCCAGGACACTTTTTC	2220
QY	2811	TTCCCTTATGATAGGCTCTCTCTCTCCACACACACACACACACACACACACACACACA	2870
Db	2221	TTCCCTTATGATAGGCTCTCTCTCTCCACACACACACACACACACACACACACACACA	2280
QY	2871	CACACACACACAAACACACACACCCCGCCCAACCAAGGTGCATGTAAAAAGATGTAGATTTC	2930
Db	2281	CACACACACACAAACACACACACCCCGCCCAACCAAGGTGCATGTAAAAAGATGTAGATTTC	2340
QY	2931	CTCTGCTTTTCTCATCTACACGCCCAGGAGGTAAAGTTAAATATAAGAGGGATTTATTGG	2990
Db	2341	CTCTGCTTTTCTCATCTACACGCCCAGGAGGTAAAGTTAAATATAAGAGGGATTTATTGG	2400
QY	2991	TAAGAGATGATGCTTAATCTGTTTAACTGTGGCCTCAAGAGAGAAATTTCTTTCTTCT	3050
Db	2401	TAAGAGATGATGCTTAATCTGTTTAACTGTGGCCTCAAGAGAGAAATTTCTTTCTTCT	2460
QY	3051	GTACTTATTTAAGCACCATTATGTGTGTAGCTTATATATACAAAGGTTTATTATATGCTA	3110
Db	2461	GTACTTATTTAAGCACCATTATGTGTGTAGCTTATATATACAAAGGTTTATTATATGCTA	2520
QY	3111	ATATAGTAATAGTAATGKTGGTGGTGTGATGATGATTAATACCATAAAAAATTAATATCCTTT	3170
Db	2521	ATATAGTAATAGTAATGKTGGTGGTGTGATGATGATTAATACCATAAAAAATTAATATCCTTT	2580
QY	3171	TAAATAAGCTAAATTAATTAATGGATCTTTTATAGTATTCATTTATGTTTTTATGTTT	3230
Db	2581	TAAATAAGCTAAATTAATTAATGGATCTTTTATAGTATTCATTTATGTTTTTATGTTT	2640
QY	3231	TTGATTTTTTAAAGACAATCTCACCTCTGTACCAGGCTGGAGTGCAGTGGTGAATCA	3290
Db	2641	TTGATTTTTTAAAGACAATCTCACCTCTGTACCAGGCTGGAGTGCAGTGGTGAATCA	2700

QY	3291	TAGCTTTTCTCAGTCTTTGAACCTCTGGGCTCAAGCAATCCTCTGCTTGGCCTGCCCTCCAAA	3350
Db	2701	TAGCTTTTCTCAGTCTTTGAACCTCTGGGCTCAAGCAATCCTCTGCTTGGCCTGCCCTCCAAA	2760
QY	3351	GTGTTGGGATACAGTCTATGAGCCACTGCATCTGGCCTAGGATCCATTTAGATTTAAATAT	3410
Db	2761	GTGTTGGGATACAGTCTATGAGCCACTGCATCTGGCCTAGGATCCATTTAGATTTAAATAT	2820
QY	3411	GCATTTTAAATTTTAAATAAATATGCTAAATTTTACCTTTATGTATGTATCTACTGGTA	3470
Db	2821	GCATTTTAAATTTTAAATAAATATGCTAAATTTTACCTTTATGTATGTATCTACTGGTA	2880
QY	3471	ATAAATCTAGTTTGTGCTTAAAGTTTAAAGTCTTTTCCAAATAGCTTTCATGTACGTGAG	3530
Db	2881	ATAAATCTAGTTTGTGCTTAAAGTTTAAAGTCTTTTCCAAATAGCTTTCATGTACGTGAG	2940
QY	3531	GGGAGACATTTAAAGTGAACACAGACAGCCAGGTGTGGTGGCTCACGCCCTGTATCCAGC	3590
Db	2941	GGGAGACATTTAAAGTGAACACAGACAGCCAGGTGTGGTGGCTCACGCCCTGTATCCAGC	3000
QY	3591	ACTCTGGGAGGCTGAGGTGGGTGATCGCTTTGAGCCCTCGAGTTTCAAGACACAGCCTGAGC	3650
Db	3001	ACTCTGGGAGGCTGAGGTGGGTGATCGCTTTGAGCCCTCGAGTTTCAAGACACAGCCTGAGC	3060
QY	3651	AACATGGCAAAACCCCTGTTTCTATAACAAAAATTTAGCCGGGCATGTGGCATGTGCCTGT	3710
Db	3061	AACATGGCAAAACCCCTGTTTCTATAACAAAAATTTAGCCGGGCATGTGGCATGTGCCTGT	3120
QY	3711	GGTCCCAGCTACTAGGGGCTGAGGAGGAGATCTTTTGGAGCCAGGAGGTCAAGGCTG	3770
Db	3121	GGTCCCAGCTACTAGGGGCTGAGGAGGAGATCTTTTGGAGCCAGGAGGTCAAGGCTG	3180
QY	3771	CACAGCAGTGTCTCGCCACTGCACCTCCAGCCTGGGTGACAGGACACAGACCTTCGCCCTC	3830
Db	3181	CACAGCAGTGTCTCGCCACTGCACCTCCAGCCTGGGTGACAGGACACAGACCTTCGCCCTC	3240
QY	3831	AAAAAATTAAGAGAAAAATTTAAAAATAAATGGAACAACTACAAAGAGCTGTGTGCTTA	3890
Db	3241	AAAAAATTAAGAGAAAAATTTAAAAATAAATGGAACAACTACAAAGAGCTGTGTGCTTA	3300
QY	3891	GATGAGCTACTTAGTTAGGCTGATATTTTGGTATTTTAACTTTTAAAGTCAGGCTGTGCTCA	3950
Db	3301	GATGAGCTACTTAGTTAGGCTGATATTTTGGTATTTTAACTTTTAAAGTCAGGCTGTGCTCA	3360
QY	3951	CCTGCACTACATTTTAAAAATATCAATTTCTCAATGTATATCCACACAAAGACTTGGTACGT	4010
Db	3361	CCTGCACTACATTTTAAAAATATCAATTTCTCAATGTATATCCACACAAAGACTTGGTACGT	3420
QY	4011	GAATGTTCACTAGTACCTTTTATTCACAAAACCCCAAGTAGAGACTATCCAAATATCCATC	4070
Db	3421	GAATGTTCACTAGTACCTTTTATTCACAAAACCCCAAGTAGAGACTATCCAAATATCCATC	3480
QY	4071	AACAAGTGAACAAATAACAAAAATGTCTATATCCATGCANTGGAATACCCCTCCAGT	4130
Db	3481	AACAAGTGAACAAATAACAAAAATGTCTATATCCATGCANTGGAATACCCCTCCAGT	3540
QY	4131	ACAAAGGAAGAAGCTACTTTGGGATCAATCCCAAGTCAATGACGCTTAAATGAAAGAGTCA	4190
Db	3541	ACAAAGGAAGAAGCTACTTTGGGATCAATCCCAAGTCAATGACGCTTAAATGAAAGAGTCA	3600
QY	4191	GACATGAAGGAGAGATAATGTATGCCATACGAAATCTTAGAAAAATGAAAGTAACTTATA	4250
Db	3601	GACATGAAGGAGAGATAATGTATGCCATACGAAATCTTAGAAAAATGAAAGTAACTTATA	3660
QY	4251	GTTACAGAAAGCAAAATCAGGGCAGGCATAGAGGCTCACACCTGTATATCCAGACCTTTGA	4310
Db	3661	GTTACAGAAAGCAAAATCAGGGCAGGCATAGAGGCTCACACCTGTATATCCAGACCTTTGA	3720
QY	4311	GAGGCCACCTGGGAAGATTTGCTAGAATCTCAGGAGTTTCAAGACACAGCCTGGGCAACACGT	4370
Db	3721	GAGGCCACCTGGGAAGATTTGCTAGAATCTCAGGAGTTTCAAGACACAGCCTGGGCAACACGT	3780
QY	4371	GAACTCCATTTCTCCACAAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTTGTCTGTGG	4430

0.

CC treatment of B cell associated immune system disorders, immunodeficiency
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency
 CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen
 CC disease, DiGeorge disease, ataxia telangiectasia, common variable
 CC immunodeficiency disorder, MHC (major histocompatibility class) class
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
 CC IgE disorder, and IgG subclass selection disorder. The DNA sequences
 CC encoding AID may be used for gene therapy and the antibodies to the AID
 CC protein may be used for diagnosis and treatment of these disorders.
 XX

SQ Sequence 2818 BP; 868 A; 548 C; 626 G; 776 T; 0 other;

Query Match 1.6%; Score 87; DB 21; Length 2818;

Best Local Similarity 100.0%; Pred. No. 1.1e-21;

Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1032 AGAGAACCATCATTAATTGAAGTGAGATTTTCTGCCTGAGACTTGCAGGGAGCAAGA 1091
 |||||

Db 1 AGAGAACCATCATTAATTGAAGTGAGATTTTCTGCCTGAGACTTGCAGGGAGCAAGA 60
 |||||

QY 1092 AGACACTCTGGACACCATATGCACAG 1118
 |||||

Db 61 AGACACTCTGGACACCATATGCACAG 87

RESULT 6

ABK52229/C

ID ABK52229 standard; cDNA; 12718 BP.

XX AC ABK52229;

XX 13-AUG-2002 (first entry)

DT cDNA encoding human CYP27A1 protein.

DE Human; Cytochrome P450; Subfamily XXVIIA; single nucleotide polymorphism;
 KW Steroid 27-Hydroxylase; Cerebrotendinous Xanthomatosis Polypeptide 1;
 KW CYP27A1; SNP; drug screening; cerebrotendinous xanthomatosis;
 KW chromosome 2q33-qter; gene; ss.
 XX Homo sapiens.

OS

XX Key Location/Qualifiers

FH variation replace (564,C)

FT /tag= a

FT /standard_name= "Single Nucleotide Polymorphism (SNP)"

FT /note= "Polymorphic site 1 (PS1)"

FT 1001..1255

FT /tag= b

FT /number= 1

FT 1001..0719

FT /tag= c

FT /product= "Human CYP27A1 protein"

FT 1256..3409

FT /tag= d

FT /number= 1

FT replace (1342,A)

FT /tag= e

FT /standard_name= "Single Nucleotide Polymorphism (SNP)"

FT /note= "Polymorphic site 2 (PS2)"

FT 3410..3600

FT /tag= f

FT /number= 2

FT replace (3564,A)

FT /tag= g

FT /standard_name= "Single Nucleotide Polymorphism (SNP)"

FT /note= "Polymorphic site 3 (PS3)"

FT replace (3592,A)

FT /tag= h

FT /standard_name= "Single Nucleotide Polymorphism (SNP)"

FT /note= "Polymorphic site 4 (PS4)"

FT 3601..5910

FT variation
 FT /number= 2
 FT /tag= j
 FT /standard_name= "Single Nucleotide Polymorphism (SNP)"
 FT /note= "Polymorphic site 5 (PS5)"
 FT replace (5770,G)
 FT /tag= k
 FT /standard_name= "Single Nucleotide Polymorphism (SNP)"
 FT /note= "Polymorphic site 6 (PS6)"
 FT 5911..6110
 FT /tag= l
 FT /number= 3
 FT replace (5970,T)
 FT /tag= m
 FT /standard_name= "Single Nucleotide Polymorphism (SNP)"
 FT /note= "Polymorphic site 7 (PS7)"
 FT replace (5988,T)
 FT /tag= n
 FT /standard_name= "Single Nucleotide Polymorphism (SNP)"
 FT /note= "Polymorphic site 8 (PS8)"
 FT 6111..6240
 FT /tag= o
 FT /number= 3
 FT 6241..6438
 FT /tag= p
 FT /number= 4
 FT 6439..6612
 FT /tag= q
 FT /number= 4
 FT 6613..6785
 FT /tag= r
 FT /number= 5
 FT replace (6689,A)
 FT /tag= s
 FT /standard_name= "Single Nucleotide Polymorphism (SNP)"
 FT /note= "Polymorphic site 9 (PS9)"
 FT 6786..7709
 FT /tag= t
 FT /number= 5
 FT replace (6843,T)
 FT /tag= u
 FT /standard_name= "Single Nucleotide Polymorphism (SNP)"
 FT /note= "Polymorphic site 10 (PS10)"
 FT 7710..7876
 FT /tag= v
 FT /number= 6
 FT replace (7843,T)
 FT /tag= w
 FT /standard_name= "Single Nucleotide Polymorphism (SNP)"
 FT /note= "Polymorphic site 11 (PS11)"
 FT 7877..8068
 FT /tag= x
 FT /number= 6
 FT replace (7931,C)
 FT /tag= y
 FT /standard_name= "Single Nucleotide Polymorphism (SNP)"
 FT /note= "Polymorphic site 12 (PS12)"
 FT replace (8003,C)
 FT /tag= z
 FT /standard_name= "Single Nucleotide Polymorphism (SNP)"
 FT /note= "Polymorphic site 13 (PS13)"
 FT 8069..8147
 FT /tag= aa
 FT /number= 7
 FT 8148..8233
 FT /tag= ab
 FT /number= 7
 FT 8234..8446
 FT /tag= ac
 FT /number= 8
 FT replace (8441,T)
 FT /tag= ad

XX 21-NOV-2001 (first entry)
XX Human reproductive system related antigen DNA SEQ ID NO: 9880.
DE Human reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
KW Homo sapiens.
XX WO200155320-A2.
PN 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01339.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.

```
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465570/50.
XX
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
XX Disclosure; SEQ ID NO 9880; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
XX Sequence 84 BP; 28 A; 20 C; 19 G; 17 T; 0 other;
SQ
Query Match 1.2%; Score 67; DB 22; Length 84;
Best Local Similarity 100.0%; Pred. No. 2.8e-14;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5446 TTGTATTATTTAGATAGAGATGGGTTTACCATGTGGCCAGGCTGGTCTCAAACTCCTG 5505
Db 73 TTGTATTATTTAGATAGATGGGTTTACCATGTGGCCAGGCTGGTCTCAAACTCCTG 14
QY 5506 ACCTCAG 5512
Db 13 ACCTCAG 7
RESULT 9
ID ABA18401
XX ID ABA18401 standard; DNA; 95 BP.
XX AC ABA18401;
XX
XX 23-JAN-2002 (first entry)
XX Human nervous system related polynucleotide SEQ ID NO 10732.
XX
XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antiskickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebrotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
XX Homo sapiens.
XX
XX WO200159063-A2.
XX
XX 16-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01334.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
```

PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246529.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249266.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX
DR WPI; 2001-541565/60.
XX
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
XX
PS Disclosure; SEQ ID NO 10732; 1701pp + Sequence Listing; English.
XX
XX

CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 95 BP; 19 A; 26 C; 22 G; 28 T; 0 other;

Query Match 1.2%; Score 67; DB 22; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5446 TTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGCTCAAACTCCTG 5505
|||||
Db 18 TTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGCTCAAACTCCTG 77
|||||

Qy 5506 ACCTCAG 5512
|||||
Db 78 ACCTCAG 84
|||||

RESULT 10
ABAL4482/C
ID ABAL4482 standard; DNA; 121 BP.
XX
AC ABAL4482;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 6813.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 03-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 03-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-541565/60.

Nucleic acids encoding 3224 human nervous system antigen polypeptides, useful for preventing, diagnosing and/or treating nervous system cancers and metastases -

Disclosure: SEQ ID NO 6813; 170lpp + Sequence Listing; English.

The invention relates to novel genes (AB11004-ABA21534) and proteins (AB114678-AB18001) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (antagonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 121 BP; 34 A; 33 C; 31 G; 23 T; 0 other;

Qy Query Match 1.28; Score 67; DB 22; Length 121;
Best Local Similarity 100.0%; Pred. No. 2.6e-14;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
5446 TTTCTATTATTAGATAGAGATGGGTTTCACCATGTTGCCAGGCTGCTCAAACTCCTG 5505
|||||
Db 79 TTTCTATTATTAGATAGAGATGGGTTTCACCATGTTGCCAGGCTGCTCAAACTCCTG 20
|||||

Qy 5506 ACCTCAG 5512

|||||
Db 19 ACCTCAG 13

RESULT 11

AAK80307/c

ID AAK80307 standard; DNA; 132 BP.

XX AAK80307;

AC AAK80307;

DT 07-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:35119.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.

OS Homo sapiens.

XX WO200157182-A2.

PN 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225269.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-483426/52.

XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -

XX Disclosure; SEQ ID NO 35119; 3071pp + Sequence Listing; English.

XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.

XX Sequence 132 BP; 36 A; 36 C; 37 G; 23 T; 0 other;

Query Match 1.2%; Score 67; DB 22; Length 132;
Best Local Similarity 100.0%; Pred. No. 2.6e-14;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5446 TTGTATTATTTAGTAGAGATGGGGTTTACCATGTTGGCAGCGTGGTCTCAAACTCCG 5505
DB 129 TTTGTATTATTTAGTAGAGATGGGGTTTACCATGTTGGCAGCGTGGTCTCAAACTCCG 70
QY 5506 ACCTCAG 5512
|||||||

Db 69 ACCTCAG 63
RESULT 12
AAK66840/C
ID AAK66840 standard; DNA; 141 BP.
XX
AC AAK66840;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/hematopoietic antigen genomic sequence SEQ ID NO:21652.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 28-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0234999.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
DR
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 21652; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 141 BP; 40 A; 34 C; 38 G; 29 T; 0 other;
Query Match 1.2%; Score 67; DB 22; Length 141;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5446 TTTGTATTTTATAGATAGATGGGTTTACACATGTTGGCCAGCTGCTCAACCTCCTG 5505
|||||
Db 106 TTTGTATTTTATAGATAGATGGGTTTACACATGTTGGCCAGCTGCTCAACCTCCTG 47
QY 5506 ACCTCAG 5512
|||||
Db 46 ACCTCAG 40
RESULT 13
AAK68098
ID AAK68098 standard; DNA; 141 BP.
XX
AC AAK68098;
XX
DT 06-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:22910.
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX


```
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX Nucleic acids encoding human immune/haematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX Disclosure; SEQ ID NO 22910; 3071pp + Sequence Listing; English.
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
XX Sequence 141 BP; 29 A; 38 C; 34 G; 40 T; 0 other;

Query Match 1.2%; Score 67; DB 22; Length 141;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5446 TTTCTATTATTAGATGGGCTTTCACCATGTTGCCAGGCTGCTCAAACTCCTG 5505
Db |||||||
Qy 5506 ACCTCAG 5512
Db |||||||
96 ACCTCAG 102

RESULT 14
AAK68878/c
ID AAK68878 standard; DNA; 141 BP.
XX AAK68878;
XX 06-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23690.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
XX 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US01354.
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225271.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
```

PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0249208.
PR 08-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PR
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis.
XX
XX
XX Disclosure; SEQ ID NO 23690; 3071pp + Sequence Listing; English.
PS
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For

CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
XX SQ Sequence 141 BP; 40 A; 34 C; 38 G; 29 T; 0 other;
Query Match 1.2%; Score 67; DB 22; Length 141;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5446 TTTGTATTTTGTAGATGGGGTTTTCACCATGTTGGCCAGGCTGCTCAAACTCCTG 5505
Db 106 TTTGTATTTTGTAGATGGGGTTTTCACCATGTTGGCCAGGCTGCTCAAACTCCTG 47
QY 5506 ACCTCAG 5512
Db 46 ACCTCAG 40
RESULT 15
AAK70059/C
ID AAK70059 standard; DNA; 141 BP.
XX
XX AAK70059;
XX
XX 06-NOV-2001 (first entry)
XX Human immune/hematopoietic antigen genomic sequence SEQ ID NO:24871.
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184564.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225216.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.

PR	14-AUG-2000;	2000US-0225268.	PR	08-NOV-2000;	2000US-0246527.
PR	14-AUG-2000;	2000US-0225270.	PR	08-NOV-2000;	2000US-0246528.
PR	14-AUG-2000;	2000US-0225447.	PR	08-NOV-2000;	2000US-0246532.
PR	14-AUG-2000;	2000US-0225757.	PR	08-NOV-2000;	2000US-0246609.
PR	14-AUG-2000;	2000US-0225758.	PR	08-NOV-2000;	2000US-0246610.
PR	14-AUG-2000;	2000US-0225759.	PR	08-NOV-2000;	2000US-0246611.
PR	18-AUG-2000;	2000US-0226279.	PR	08-NOV-2000;	2000US-0246613.
PR	22-AUG-2000;	2000US-0226681.	PR	08-NOV-2000;	2000US-0249207.
PR	22-AUG-2000;	2000US-0226868.	PR	17-NOV-2000;	2000US-0249208.
PR	22-AUG-2000;	2000US-0227182.	PR	17-NOV-2000;	2000US-0249209.
PR	23-AUG-2000;	2000US-0227009.	PR	17-NOV-2000;	2000US-0249210.
PR	30-AUG-2000;	2000US-0228924.	PR	17-NOV-2000;	2000US-0249211.
PR	01-SEP-2000;	2000US-0229287.	PR	17-NOV-2000;	2000US-0249212.
PR	01-SEP-2000;	2000US-0229343.	PR	17-NOV-2000;	2000US-0249213.
PR	01-SEP-2000;	2000US-0229344.	PR	17-NOV-2000;	2000US-0249214.
PR	01-SEP-2000;	2000US-0229345.	PR	17-NOV-2000;	2000US-0249215.
PR	05-SEP-2000;	2000US-0229509.	PR	17-NOV-2000;	2000US-0249216.
PR	05-SEP-2000;	2000US-0229513.	PR	17-NOV-2000;	2000US-0249217.
PR	06-SEP-2000;	2000US-0230437.	PR	17-NOV-2000;	2000US-0249218.
PR	06-SEP-2000;	2000US-0230438.	PR	17-NOV-2000;	2000US-0249244.
PR	08-SEP-2000;	2000US-0231242.	PR	17-NOV-2000;	2000US-0249245.
PR	08-SEP-2000;	2000US-0231243.	PR	17-NOV-2000;	2000US-0249264.
PR	08-SEP-2000;	2000US-0231244.	PR	17-NOV-2000;	2000US-0249265.
PR	08-SEP-2000;	2000US-0231413.	PR	17-NOV-2000;	2000US-0249297.
PR	08-SEP-2000;	2000US-0232080.	PR	17-NOV-2000;	2000US-0249300.
PR	08-SEP-2000;	2000US-0232081.	PR	01-DEC-2000;	2000US-0250160.
PR	12-SEP-2000;	2000US-0231968.	PR	01-DEC-2000;	2000US-0250391.
PR	14-SEP-2000;	2000US-0232397.	PR	05-DEC-2000;	2000US-0251030.
PR	14-SEP-2000;	2000US-0232398.	PR	05-DEC-2000;	2000US-0251988.
PR	14-SEP-2000;	2000US-0232399.	PR	05-DEC-2000;	2000US-0256719.
PR	14-SEP-2000;	2000US-0232400.	PR	06-DEC-2000;	2000US-0251479.
PR	14-SEP-2000;	2000US-0232401.	PR	08-DEC-2000;	2000US-0251856.
PR	14-SEP-2000;	2000US-0233063.	PR	08-DEC-2000;	2000US-0251856.
PR	14-SEP-2000;	2000US-0233064.	PR	08-DEC-2000;	2000US-0251868.
PR	14-SEP-2000;	2000US-0233065.	PR	08-DEC-2000;	2000US-0251989.
PR	21-SEP-2000;	2000US-0234223.	PR	08-DEC-2000;	2000US-0251990.
PR	21-SEP-2000;	2000US-0234274.	PR	11-DEC-2000;	2000US-0254097.
PR	25-SEP-2000;	2000US-0234997.	PR	05-JAN-2001;	2001US-0259678.
PR	25-SEP-2000;	2000US-0234998.	XX		
PR	26-SEP-2000;	2000US-0235484.	PA	(HUMA-) HUMAN GENOME SCI INC.	
PR	27-SEP-2000;	2000US-0235834.	XX		
PR	27-SEP-2000;	2000US-0235836.	PI	Rosen CA, Barash SC, Ruben SM;	
PR	29-SEP-2000;	2000US-0236327.	XX	WPI; 2001-483426/52.	
PR	29-SEP-2000;	2000US-0236367.	DR		
PR	29-SEP-2000;	2000US-0236368.	XX		
PR	29-SEP-2000;	2000US-0236369.	PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,	
PR	29-SEP-2000;	2000US-0236370.	PT	useful for preventing, diagnosing and/or treating cancers and	
PR	02-OCT-2000;	2000US-0236802.	XX	metastasis -	
PR	02-OCT-2000;	2000US-0237037.	XX		
PR	02-OCT-2000;	2000US-0237038.	PS	Disclosure; SEQ ID NO 24871; 3071pp + Sequence Listing; English.	
PR	02-OCT-2000;	2000US-0237039.	XX		
PR	13-OCT-2000;	2000US-0237040.	CC	AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)	
PR	13-OCT-2000;	2000US-0239935.	CC	amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic	
PR	20-OCT-2000;	2000US-0239937.	CC	activity, and can be used in gene therapy and vaccine production. (I)	
PR	20-OCT-2000;	2000US-0240960.	CC	proteins and polynucleotides may be used in the prevention, diagnosis and	
PR	20-OCT-2000;	2000US-0241221.	CC	treatment of diseases associated with inappropriate (I) expression. For	
PR	20-OCT-2000;	2000US-0241785.	CC	example, they may be used to treat disorders associated with decreased	
PR	20-OCT-2000;	2000US-0241786.	CC	expression by rectifying mutations or deletions in a patient's genome	
PR	20-OCT-2000;	2000US-0241787.	CC	that affect the activity of (I) by expressing inactive proteins or to	
PR	20-OCT-2000;	2000US-0241808.	CC	supplement the patients own production of (I). Additionally, (I)	
PR	20-OCT-2000;	2000US-0241809.	CC	polynucleotides may be used to produce the secreted (I), by inserting	
PR	20-OCT-2000;	2000US-0241826.	CC	the nucleic acids into a host cell and culturing the cell to express the	
PR	01-NOV-2000;	2000US-0244617.	CC	protein. (I) proteins and polynucleotides may be used to prevent,	
PR	08-NOV-2000;	2000US-0246474.	CC	diagnose and treat immune/haematopoietic-related diseases, especially	
PR	08-NOV-2000;	2000US-0246475.	CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703	
PR	08-NOV-2000;	2000US-0246476.	CC	to AAK87694 represent human immune/haematopoietic antigen genomic	
PR	08-NOV-2000;	2000US-0246477.	CC	sequences from the present invention. AAK54942 to AAK54950 and AAK82169	
PR	08-NOV-2000;	2000US-0246478.	CC	represent sequences used in the exemplification of the present invention.	
PR	08-NOV-2000;	2000US-0246523.	XX		
PR	08-NOV-2000;	2000US-0246524.	SQ	Sequence 141 BP; 40 A; 34 C; 38 G; 29 T; 0 other;	
PR	08-NOV-2000;	2000US-0246525.			
PR	08-NOV-2000;	2000US-0246526.			

Query Match

1.2%; Score 67; DB 22; Length 141;

```

Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5446 TTTGTATTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCCTG 5505
Db 106 TTTGTATTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCCTG 47
QY 5506 ACCTCAG 5512
Db 46 ACCTCAG 40

```

Search completed: June 19, 2003, 11:52:01
Job time : 1074 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 11:34:27 ; Search time 720 seconds
(without alignments)
11238.046 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 5514
Sequence: 1 acagagacacatggtccca.....tcaaaactcctgacctcagag 5514

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1042519 seqs, 733713590 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5514	100.0	5514	9	US-09-966-880A-9
2	4771	86.5	11204	9	US-09-966-880A-35
3	87	1.6	87	9	US-09-966-880A-11
4	87	1.6	2818	9	US-09-966-880A-7
5	68	1.2	8158	9	US-09-832-232-30
c 6	67	1.2	84	9	US-09-764-891-9880
7	67	1.2	1160	9	US-09-764-891-5684
c 8	67	1.2	5930	10	US-09-764-877-2384
9	67	1.2	11754	9	US-09-984-827-5
c 10	67	1.2	16086	10	US-09-764-877-2385
c 11	67	1.2	24028	9	US-09-764-891-8207
c 12	67	1.2	32170	9	US-10-074-095-1108
c 13	67	1.2	32170	10	US-09-764-860-1108
c 14	67	1.2	34668	9	US-09-900-449A-3
c 15	67	1.2	59837	10	US-09-982-091A-5
c 16	67	1.2	122186	9	US-09-563-728A-36
c 17	67	1.2	143306	10	US-09-729-920A-3
c 18	67	1.2	174566	9	US-10-020-141-1
19	66	1.2	251	9	US-10-092-154-1205

20	66	1.2	251	10	US-09-764-847-1205
21	66	1.2	285	9	US-10-091-504-1783
22	66	1.2	285	10	US-09-764-869-1783
23	66	1.2	638	9	US-10-091-572-628
24	66	1.2	638	9	US-09-764-891-6781
25	66	1.2	6257	9	US-10-091-504-1782
26	66	1.2	6257	10	US-09-764-869-1782
27	66	1.2	6761	9	US-10-091-504-1724
28	66	1.2	6761	10	US-09-764-869-1724
c 29	66	1.2	7219	9	US-10-102-627-107
30	66	1.2	8758	10	US-09-962-276-3
31	66	1.2	12118	10	US-09-788-654A-3
32	66	1.2	14093	9	US-10-092-154-1744
33	66	1.2	14093	10	US-09-764-847-1744
34	66	1.2	28588	9	US-10-073-961-399
35	66	1.2	28588	10	US-09-764-887-399
36	66	1.2	32146	9	US-10-074-095-797
37	66	1.2	32146	10	US-09-764-860-797
38	66	1.2	32248	9	US-10-074-095-802
39	66	1.2	32248	10	US-09-764-860-802
40	66	1.2	52216	10	US-09-747-810-1
41	65	1.2	110096	10	US-09-880-107-1542
c 42	64	1.2	426	9	US-09-918-995-4335
c 43	64	1.2	3569	9	US-10-091-572-737
c 44	64	1.2	3569	9	US-09-764-891-9161
45	64	1.2	10527	9	US-10-116-420-100

ALIGNMENTS

RESULT 1

US-09-966-880A-9
; Sequence 9, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasaku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JF00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 5514
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: intron
; LOCATION: (1)...(1031)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1032)...(1118)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (1119)...(5514)
; US-09-966-880A-9

Query Match 100.0%; Score 5514; DB 9; Length 5514;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ACAGACGAATACATGGTCCAAAGCTAGGCTATTGATTTGAAATCATCAAGGTATAGATG 60
|||||

Db 1 ACAGACGAATACATGGTCCAAAGCTAGGGCTATTGATTTGAAATCATCAAGGTATAGATG 60
QY 61 GTATCAAAAGGCTTGAGGCAGGAAGAGAGACAGAGACCCCTAGCTGCATTGCTTAGCATTTGCA 120
Db 61 GTATCAAAAGGCTTGAGGCAGGAAGAGAGAGACAGAGACCCCTAGCTGCATTGCTTAGCATTTGCA 120
QY 121 TCCCTAGCACCTGGCATAGTTTCCCAATTAAACAGTAGGCATGAAGTATCTACTCAGTGAATA 180
Db 121 TCCCTAGCACCTGGCATAGTTTCCCAATTAAACAGTAGGCATGAAGTATCTACTCAGTGAATA 180
QY 181 AATAGAAATGCATATGGGCTACAGTAGGAGAGAGAAATAAAATCTTTAATAGACCAGATTC 240
Db 181 AATAGAAATGCATATGGGCTACAGTAGGAGAGAGAAATAAAATCTTTAATAGACCAGATTC 240
QY 241 TATGAGAGACAAAAATTAAGTCTTTTATTTCGAAGATCTTAGCCCTGTTTCCCAATTCAG 300
Db 241 TATGAGAGACAAAAATTAAGTCTTTTATTTCGAAGATCTTAGCCCTGTTTCCCAATTCAG 300
QY 301 TGCAGCCAGTTAGACACATGATCTGTCTGGTGAACAAGCAATTTTGTATTTTGGGGGAC 360
Db 301 TGCAGCCAGTTAGACACATGATCTGTCTGGTGAACAAGCAATTTTGTATTTTGGGGGAC 360
QY 361 TCGTCTGCTTCTGACTCCAAATTAAGGATTTTTTTTTTCTAAAAAGATGGCTCATG 420
Db 361 TCGTCTGCTTCTGACTCCAAATTAAGGATTTTTTTTTTCTAAAAAGATGGCTCATG 420
QY 421 CAAAAATCACTCTTTGGTCTAAATATCTAGTCTTCAAGCAATCTTGTATGCAATCAGA 480
Db 421 CAAAAATCACTCTTTGGTCTAAATATCTAGTCTTCAAGCAATCTTGTATGCAATCAGA 480
QY 481 AAGAAAAAATCCATGGTTTGGAGGCAAAAATTTTGTCTCTAAATCTATATAACTGA 540
Db 481 AAGAAAAAATCCATGGTTTGGAGGCAAAAATTTTGTCTCTAAATCTATATAACTGA 540
QY 541 GTTCATTTCCTTAAGTCAAGACAGGAGCTGTAGTGCCTGTCTGACTGAGGTTTCAGAG 600
Db 541 GTTCATTTCCTTAAGTCAAGACAGGAGCTGTAGTGCCTGTCTGACTGAGGTTTCAGAG 600
QY 601 AGACTGTGGGAATATGGGGGAATTAAGAGCTATCTGAGGCTCTTCAACACAATAAAGCCAA 660
Db 601 AGACTGTGGGAATATGGGGGAATTAAGAGCTATCTGAGGCTCTTCAACACAATAAAGCCAA 660
QY 661 GAAGCTATTAAATGCTCTTTAAGTATTTACATAAATATTAATCTCAATTTGCTCTTT 720
Db 661 GAAGCTATTAAATGCTCTTTAAGTATTTACATAAATATTAATCTCAATTTGCTCTTT 720
QY 721 TATTTTGTCTATCATGATTATAATTGAAGTGTCTACTCTTACTGCTCTCTGATCTTTGC 780
Db 721 TATTTTGTCTATCATGATTATAATTGAAGTGTCTACTGTGTACTGCTCTCTGATCTTTGC 780
QY 781 TAGCTATGAGCATGGACTGGGCTTTTAGAGCAGACGCCCAAGGAACCTTAAACATTAA 840
Db 781 TAGCTATGAGCATGGACTGGGCTTTTAGAGCAGACGCCCAAGGAACCTTAAACATTAA 840
QY 841 AGCAGAGTGCCTCAATGGTTTAACTGTGTGACTCTGCCTATGACAGCCCAAGCCCAAGCC 900
Db 841 AGCAGAGTGCCTCAATGGTTTAACTGTGTGACTCTGCCTATGACAGCCCAAGCCCAAGCC 900
QY 901 CATCTTCACTGGATCCAAATCAGGACAGGCGGTTGGGATACCTGGTGGGGTGATGCT 960
Db 901 CATCTTCACTGGATCCAAATCAGGACAGGCGGTTGGGATACCTGGTGGGGTGATGCT 960
QY 961 GTCAGGGAGGAGCCCAAGGCAAGCTCAAAATTTGAATGTGAAGGGCAATGCACCTGT 1020
Db 961 GTCAGGGAGGAGCCCAAGGCAAGCTCAAAATTTGAATGTGAAGGGCAATGCACCTGT 1020
QY 1021 CAGACTGACAGACAGAACCATCAATTAATTGAAGTGAGATTTTCTGGCTGAGACTTGCA 1080
Db 1021 CAGACTGACAGACAGAACCATCAATTAATTGAAGTGAGATTTTCTGGCTGAGACTTGCA 1080
QY 1081 GGGAGCAAGAGACACTCTGGACACCCTATGGACAGGTAAAGAGGCGAGTCTTCTCGTG 1140
Db 1081 GGGAGCAAGAGACACTCTGGACACCCTATGGACAGGTAAAGAGGCGAGTCTTCTCGTG 1140

QY 1141 GGTGATTGCACCTGGCCTTCCTCTCAGAGCAAAATCTGAGTAATGAGACTGCTAGCTATGCC 1200
Db 1141 GGTGATTGCACCTGGCCTTCCTCTCAGAGCAAAATCTGAGTAATGAGACTGCTAGCTATGCC 1200
QY 1201 TTTTCTCTCATGTAACTGTCTGACTGATAGATCAGCTTGATCAATATGATATATATTTT 1260
Db 1201 TTTTCTCTCATGTAACTGTCTGACTGATAGATCAGCTTGATCAATATGATATATATTTT 1260
QY 1261 TTGATCTGTCTCCTTTTCTTCTTATTCAGATCTTATACGTGTGAGCCCAATTTCTTCTGT 1320
Db 1261 TTGATCTGTCTCCTTTTCTTCTTATTCAGATCTTATACGTGTGAGCCCAATTTCTTCTGT 1320
QY 1321 TTCAGACTTCTCTTGATTTCCCTCTTTTTCATGTGGCAAAAGAGTAGTGCCTGACATGT 1380
Db 1321 TTCAGACTTCTCTTGATTTCCCTCTTTTTCATGTGGCAAAAGAGTAGTGCCTGACATGT 1380
QY 1381 ACTGATTCGCTCTGAGATTTGTACCATGGTTGAACTAAATTTATGCTAAATATTAACA 1440
Db 1381 ACTGATTCGCTCTGAGATTTGTACCATGGTTGAACTAAATTTATGCTAAATATTAACA 1440
QY 1441 TAGCAAACTCTTTAGAGACTCAAAATCATGAAAGGTAAATAGCAGTACTGTACTAAAAACGG 1500
Db 1441 TAGCAAACTCTTTAGAGACTCAAAATCATGAAAGGTAAATAGCAGTACTGTACTAAAAACGG 1500
QY 1501 TAGTGTAAATTTTCGTAATAATTTTGTAAATATTCACAGTAAACAACTTTGAAGACACA 1560
Db 1501 TAGTGTAAATTTTCGTAATAATTTTGTAAATATTCACAGTAAACAACTTTGAAGACACA 1560
QY 1561 CTTTCTCTAGGAGGCGTTACTGAAATTAATTTAGCTATAGTAAAGAAATTTGTAATTTAG 1620
Db 1561 CTTTCTCTAGGAGGCGTTACTGAAATTAATTTAGCTATAGTAAAGAAATTTGTAATTTAG 1620
QY 1621 AAATGCCAAGCATTTTAAATTAATGCTTGAAGTCACTATGATGTGTCATTATAAGG 1680
Db 1621 AAATGCCAAGCATTTTAAATTAATGCTTGAAGTCACTATGATGTGTCATTATAAGG 1680
QY 1681 AGACAAATTCATTCAAGCAAGTTTAAATTTAAAGGCGCAATTTGTTAGGCAAGTTAATG 1740
Db 1681 AGACAAATTCATTCAAGCAAGTTTAAATTTAAAGGCGCAATTTGTTAGGCAAGTTAATG 1740
QY 1741 GCACCTTTTACTATTAACCTCAATTTCCATTTGTTTCCAGAGTACTTAACTTACCTCTTAG 1800
Db 1741 GCACCTTTTACTATTAACCTCAATTTCCATTTGTTTCCAGAGTACTTAACTTACCTCTTAG 1800
QY 1801 GTGTGAATTTGGTTAAGGCTCTCATATATGCTTTTATGTGCAGTTTTGATAGGTTATTGT 1860
Db 1801 GTGTGAATTTGGTTAAGGCTCTCATATATGCTTTTATGTGCAGTTTTGATAGGTTATTGT 1860
QY 1861 CATAGAATTTATTTCTATTTCTACATTTTACTATGATGATGTATGAGAAATACACCTA 1920
Db 1861 CATAGAATTTATTTCTATTTCTACATTTTACTATGATGATGTATGAGAAATACACCTA 1920
QY 1921 ATCCCTTATCTTTACCTCAATTTTAACTTCTTAAAGAACTTACATTTACAGAAATAAGA 1980
Db 1921 ATCCCTTATCTTTACCTCAATTTTAACTTCTTAAAGAACTTACATTTACAGAAATAAGA 1980
QY 1981 TTTTAAAAAATATATTTTTTTTGTAGACAGAGGCTTTAGCCCGAGCGGCTGCTCT 2040
Db 1981 TTTTAAAAAATATATTTTTTTTGTAGACAGAGGCTTTAGCCCGAGCGGCTGCTCT 2040
QY 2041 AAGTCTGGCCCAAGGATCCTCCTCGCTGGGCTCCTTAAAGTGTGGAATTTATAGACAT 2100
Db 2041 AAGTCTGGCCCAAGGATCCTCCTCGCTGGGCTCCTTAAAGTGTGGAATTTATAGACAT 2100
QY 2101 GAGCCATCACATCCAATATACAGAAATAAGATTTTTTAATGGAGGATTTAATGTTCTTCAG 2160
Db 2101 GAGCCATCACATCCAATATACAGAAATAAGATTTTTTAATGGAGGATTTAATGTTCTTCAG 2160
QY 2161 AAAATTTTCTTGAGGTCAGACAAATGTCTCTCCTCAGTTTACACTGAGATTTTGAA 2220
Db 2161 AAAATTTTCTTGAGGTCAGACAAATGTCTCTCCTCAGTTTACACTGAGATTTTGAA 2220

QY	2221	AACAAGCTGTAGAGCTATAGGTCCTTGTGAAGGGTCCATTTGGAAATACTTGTTCAAAGTAAA	2280
Db	2221	AACAAGCTGTAGAGCTATAGGTCCTTGTGAAGGGTCCATTTGGAAATACTTGTTCAAAGTAAA	2280
QY	2281	ATGGAAGCAAAAGGTAAAAATCAGCAGCTGGAATTTACAGAAAGACAGAAAGGAGAAAAG	2340
Db	2281	ATGGAAGCAAAAGGTAAAAATCAGCAGCTGGAATTTACAGAAAGACAGAAAGGAGAAAAG	2340
QY	2341	ATGAAATTCACAAGGACAGAAGGGAATATATTATCATTTAAGGAGGACAGTATCTGTAGA	2400
Db	2341	ATGAAATTCACAAGGACAGAAGGGAATATATTATCATTTAAGGAGGACAGTATCTGTAGA	2400
QY	2401	GCTCATTTAGTGATGGCAAAATGACTTGGTCAGAGTATTTTAAACC CGCTTGTTCCTGGT	2460
Db	2401	GCTCATTTAGTGATGGCAAAATGACTTGGTCAGAGTATTTTAAACC CGCTTGTTCCTGGT	2460
QY	2461	TTGCACGGCTGGGGATGTCAGCTAGGCTTGCCCTCAGGGAGCAGACGTGTCCAGAGCAGC	2520
Db	2461	TTGCACGGCTGGGGATGTCAGCTAGGCTTGCCCTCAGGGAGCAGACGTGTCCAGAGCAGC	2520
QY	2521	TGTCAGCCTGCAAGCCGTGAAACACTCCCTCGGTAAAGTCCCTTACTCAGGACAGAAAAT	2580
Db	2521	TGTCAGCCTGCAAGCCGTGAAACACTCCCTCGGTAAAGTCCCTTACTCAGGACAGAAAAT	2580
QY	2581	GACGAGAACAGGGAGCTGGAACAGGCCCTTAACAGAGAAAGGAAAGTAAATGATCAACA	2640
Db	2581	GACGAGAACAGGGAGCTGGAACAGGCCCTTAACAGAGAAAGGAAAGTAAATGATCAACA	2640
QY	2641	AACTTAACTAGCAGGTCAGGATCACGCAATTCATTTCACTCTCAGCTGGTAACATGTGACA	2700
Db	2641	AACTTAACTAGCAGGTCAGGATCACGCAATTCATTTCACTCTCAGCTGGTAACATGTGACA	2700
QY	2701	GAAACAGTGTAGGCTTATTGTATTTTCATGTAGAGTAGGACCCCAAAAATCCACCCAAAGT	2760
Db	2701	GAAACAGTGTAGGCTTATTGTATTTTCATGTAGAGTAGGACCCCAAAAATCCACCCAAAGT	2760
QY	2761	CCTTTATCTATGCCACATCCCTTATCTATACCTCCAGGACACTTTTCTTCCCTATAGA	2820
Db	2761	CCTTTATCTATGCCACATCCCTTATCTATACCTCCAGGACACTTTTCTTCCCTATAGA	2820
QY	2821	TAAGGCTCTCTCTCTCCA	2880
Db	2821	TAAGGCTCTCTCTCTCCA	2880
QY	2881	CACAAACACACACCCCGCCCAACCAAGGTGCATGTAAAAAGATGTAGATTCTCTGCGCTTT	2940
Db	2881	CACAAACACACACCCCGCCCAACCAAGGTGCATGTAAAAAGATGTAGATTCTCTGCGCTTT	2940
QY	2941	CTCATCTACACGCCAGGAGGTAGTTAAATATAAGAGGGATTTTATGGTAAGAGATGA	3000
Db	2941	CTCATCTACACGCCAGGAGGTAGTTAAATATAAGAGGGATTTTATGGTAAGAGATGA	3000
QY	3001	TGCTTAATCTGTTTAAACACTGGGCCCTCAAGAGAGAAATTCCTTTCTCTGTACTATTTA	3060
Db	3001	TGCTTAATCTGTTTAAACACTGGGCCCTCAAGAGAGAAATTCCTTTCTCTGTACTATTTA	3060
QY	3061	AGCACCTATTATGTGTGTAGCTTATATACAAAGGGTTATTATGCTAAATATAGTAAT	3120
Db	3061	AGCACCTATTATGTGTGTAGCTTATATACAAAGGGTTATTATGCTAAATATAGTAAT	3120
QY	3121	AGTAATGKTGGTGGTACTATGTAAATACCATAAAAAATTAATATCCTTTAAAAATAAAG	3180
Db	3121	AGTAATGKTGGTGGTACTATGTAAATACCATAAAAAATTAATATCCTTTAAAAATAAAG	3180
QY	3181	CTAAATTATATTGGATCTTTTATAGTATTCATTTTATGTTTTTATGTTTTTTCGATTTTT	3240
Db	3181	CTAAATTATATTGGATCTTTTATAGTATTCATTTTATGTTTTTATGTTTTTTCGATTTTT	3240
QY	3241	AAAAGACAATCTCACCTCTGTACCCAGGCTGGAGTGCAGTGGTGCATCATCATCTCTC	3300
Db	3241	AAAAGACAATCTCACCTCTGTACCCAGGCTGGAGTGCAGTGGTGCATCATCATCTCTC	3300
QY	3301	CAGTCTTGAACTCCTGGGCTCAAGCAATCCTCCTGCTTGGCTTCCCAAAAGTGTGGGAT	3360

	Db	3301	 CAGTCTTTGAACCTCTCGGCTCAAGCAATCTCTCGCTTGCCCTGCCAATGTGGGAT	3360
	Qy	3361	ACAGTCATGAGCCACTGCATCTGGCCTAGGATCCATTAGATTAAAAATATGCAATTTTAAA 	3420
	Db	3361	ACAGTCATGAGCCACTGCATCTGGCCTAGGATCCATTAGATTAAAAATATGCAATTTTAAA 	3420
	Qy	3421	TTTTAAATAATATGGGTAAATTTTACCTTATGTAAATGTATACCTGGTAAATAAATCTAG 	3480
	Db	3421	TTTTAAATAATATGGGTAAATTTTACCTTATGTAAATGTATACCTGGTAAATAAATCTAG 	3480
	Qy	3481	TTTCTGCTTAAAGTTTAAAGTCTTTTCCAAATAGCTTTCATGTACGTGAGGGGAGACAATT 	3540
	Db	3481	TTTCTGCTTAAAGTTTAAAGTCTTTTCCAAATAGCTTTCATGTACGTGAGGGGAGACAATT 	3540
	Qy	3541	TAAAGTGAACACAGACAGCCAGGTGTGGTGGCTCACGGCTGTAAATCCCAGACACTCTGGGAG 	3600
	Db	3541	TAAAGTGAACACAGACAGCCAGGTGTGGTGGCTCACGGCTGTAAATCCCAGACACTCTGGGAG 	3600
	Qy	3601	GCTGAGGTGGGTGGATCGCTTGGAGCCCTGGAGTTCAAAGACAGCCTGAGCAACATGGCAA 	3660
	Db	3601	GCTGAGGTGGGTGGATCGCTTGGAGCCCTGGAGTTCAAAGACAGCCTGAGCAACATGGCAA 	3660
	Qy	3661	AACCTGTCTTATTAACAAAAATTTAGCCGGGCATGGTGGCATGTGCCTGTGGTCCCCAGCT 	3720
	Db	3661	AACCTGTCTTATTAACAAAAATTTAGCCGGGCATGGTGGCATGTGCCTGTGGTCCCCAGCT 	3720
	Qy	3721	ACTAGGGGCTGAGCGAGGAGAATCTTTTGGAGCCACAGAGGTCAAGGCTGCACGTAGCAG 	3780
	Db	3721	ACTAGGGGCTGAGCGAGGAGAATCTTTTGGAGCCACAGAGGTCAAGGCTGCACGTAGCAG 	3780
	Qy	3781	TGCTTGGCCCACTGCACCTCCAGCCTGGGTGACAGGACCAGACCTTGCCCTCAAAAAATAA 	3840
	Db	3781	TGCTTGGCCCACTGCACCTCCAGCCTGGGTGACAGGACCAGACCTTGCCCTCAAAAAATAA 	3840
	Qy	3841	GAAAGAAAAATTTAAAAATTAATGGAAAAACAACACTACAAAGAGCTGTGTCTTAGATGAGCTAC 	3900
	Db	3841	GAAAGAAAAATTTAAAAATTAATGGAAAAACAACACTACAAAGAGCTGTGTCTTAGATGAGCTAC 	3900
	Qy	3901	TTAGTTAGGCTGATATTTTGGTATTTTAACTTTTAAAGTCAGGGTCTGTCACTCGCACTAC 	3960
	Db	3901	TTAGTTAGGCTGATATTTTGGTATTTTAACTTTTAAAGTCAGGGTCTGTCACTCGCACTAC 	3960
	Qy	3961	ATTATTAATAATCAATTTCAATGTATATCCACACAAAGACTGGTACGTGAATGTTTCAT 	4020
	Db	3961	ATTATTAATAATCAATTTCAATGTATATCCACACAAAGACTGGTACGTGAATGTTTCAT 	4020
	Qy	4021	AGTACCTTTATTCAAAAACCCCAAGTAGAGACTATCCAATATCCATCAACAAGTGAA 	4080
	Db	4021	AGTACCTTTATTCAAAAACCCCAAGTAGAGACTATCCAATATCCATCAACAAGTGAA 	4080
	Qy	4081	CAAAATAACAAAAATGTCTATATCCATGCAATGGAATACCACCCTGCAGTACAAAGGAAG 	4140
	Db	4081	CAAAATAACAAAAATGTCTATATCCATGCAATGGAATACCACCCTGCAGTACAAAGGAAG 	4140
	Qy	4141	AAGCTACTTGGGGATGAATCCCAAGTCATGACGCTAAATGAAAGACTCAGACATGAAGG 	4200
	Db	4141	AAGCTACTTGGGGATGAATCCCAAGTCATGACGCTAAATGAAAGACTCAGACATGAAGG 	4200
	Qy	4201	AGGAGATAATGTATGCCATACGAAATTTCTAGAAAAATGAAAGTAACTTTATAGTTACAGAAA 	4260
	Db	4201	AGGAGATAATGTATGCCATACGAAATTTCTAGAAAAATGAAAGTAACTTTATAGTTACAGAAA 	4260
	Qy	4261	GCAATCAGGCGAGGCATAGAGGCTCACCTGTAAATCCCAGCACTTTGAGAGGCCACGT 	4320
	Db	4261	GCAATCAGGCGAGGCATAGAGGCTCACCTGTAAATCCCAGCACTTTGAGAGGCCACGT 	4320
	Qy	4321	GGGAAGATTGCTAGAACCTCAGGAGTTCAAGACCACGCTGGGCAACACAGTGAATCCAT 	4380
	Db	4321	GGGAAGATTGCTAGAACCTCAGGAGTTCAAGACCACGCTGGGCAACACAGTGAATCCAT 	4380
	Qy	4381	TCTCCACAAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTTGTCTGTGGGAGGGGAAG 	4440

Db 4381 TCTCCACAAAATGGAAAAAAGAAAGCAAAATCAGTGGTTCTCTGTGGGAGGGGAG 4440

QY 4441 GACTCAAGAGAGGAGAGCTCTGTGGGGTGAGGGTGGTGAATTCAGGTTCTGTATCT 4500

Db 4441 GACTCAAGAGAGGAGAGCTCTGTGGGGTGAGGGTGGTGAATTCAGGTTCTGTATCT 4500

QY 4501 GACTGTGTAGCAGTTTGGGGTGTACATCCAAAATATTCGTAGAATATGCAATCTTA 4560

Db 4501 GACTGTGTAGCAGTTTGGGGTGTACATCCAAAATATTCGTAGAATATGCAATCTTA 4560

QY 4561 AATGGGTGAGTTTACTGTATCTAAATATATCACTCAATGTAAGAAAAAATATGCTAAG 4620

Db 4561 AATGGGTGAGTTTACTGTATCTAAATATATCACTCAATGTAAGAAAAAATATGCTAAG 4620

QY 4621 AAAAGTTTCAATCTCTGCCAGCAACGTTATTCAAATCTCGAGCCCTTTACTTCGCA 4680

Db 4621 AAAAGTTTCAATCTCTGCCAGCAACGTTATTCAAATCTCGAGCCCTTTACTTCGCA 4680

QY 4681 AATTCCTGCACTTCTGCCCTGACCATTTAGGTGACAGCACTAGCTCCACAATTTGGATA 4740

Db 4681 AATTCCTGCACTTCTGCCCTGACCATTTAGGTGACAGCACTAGCTCCACAATTTGGATA 4740

QY 4741 AATGCATTTCTGAAAAGACTAGGCAAAAATCCAGGCATCACTGTGCTTCATATCAA 4800

Db 4741 AATGCATTTCTGAAAAGACTAGGCAAAAATCCAGGCATCACTGTGCTTCATATCAA 4800

QY 4801 CCACGCTGTACAGCTTGTGCTGTGCTGCTGCTGCAATGCGGACTCTTTGATTTCTTAA 4860

Db 4801 CCACGCTGTACAGCTTGTGCTGTGCTGCTGCTGCAATGCGGACTCTTTGATTTCTTAA 4860

QY 4861 GGAACCTGGTTACAGAGTATTTCCACAATATGTTCAATTAAGTCTGCTTATGATG 4920

Db 4861 GGAACCTGGTTACAGAGTATTTCCACAATATGTTCAATTAAGTCTGCTTATGATG 4920

QY 4921 CAAGACACTGTCTAGGAGCCAGAAAACAAAGAGGAGGAATCAGTCAATATGCGGA 4980

Db 4921 CAAGACACTGTCTAGGAGCCAGAAAACAAAGAGGAGGAATCAGTCAATATGCGGA 4980

QY 4981 ACAACATGCAAGATATTTAGATCATTTTGTAGTATTAATAAGCAGCAGAGTACAAAAT 5040

Db 4981 ACAACATGCAAGATATTTAGATCATTTTGTAGTATTAATAAGCAGCAGAGTACAAAAT 5040

QY 5041 CACACATGCAATCAGTATTAATCCAAATCATGTAATATGCTGTAGAAAAGCTAGAGG 5100

Db 5041 CACACATGCAATCAGTATTAATCCAAATCATGTAATATGCTGTAGAAAAGCTAGAGG 5100

QY 5101 AATAAACACAAGAACTTAAACAGTCAATGCTCATTTAGACACTAAGTCTAATTTATTTAT 5160

Db 5101 AATAAACACAAGAACTTAAACAGTCAATGCTCATTTAGACACTAAGTCTAATTTATTTAT 5160

QY 5161 AGACACTATGATATTTAGATTTAAAAATCTTTTAAATTTTAAAAATTTAGAGCTCTCT 5220

Db 5161 AGACACTATGATATTTAGATTTAAAAATCTTTTAAATTTTAAAAATTTAGAGCTCTCT 5220

QY 5221 ATTTTTCATAGTATTCAGATTTGCAATGATCAAGTATTAATTTTCTTTTCTTTTCTTT 5280

Db 5221 ATTTTTCATAGTATTCAGATTTGCAATGATCAAGTATTAATTTTCTTTTCTTTTCTTT 5280

QY 5281 TTTTCTTTTCTTTTGTAGATGGAGTTTGGTCTTTGTTGCCCATGCTGGAGTGAATGGCATG 5340

Db 5281 TTTTCTTTTCTTTTGTAGATGGAGTTTGGTCTTTGTTGCCCATGCTGGAGTGAATGGCATG 5340

QY 5341 AYCATAAGCTCACTGCAACCTCCACCTCTCGGGTTCAAGCAAAAGCTGTCGCCCTCAGCTCC 5400

Db 5341 AYCATAAGCTCACTGCAACCTCCACCTCTCGGGTTCAAGCAAAAGCTGTCGCCCTCAGCTCC 5400

QY 5401 CGGGTAGATGGGATACAGCGCCGCCACACACCTCGGCTAATGTTTGTATTTTATGTA 5460

Db 5401 CGGGTAGATGGGATACAGCGCCGCCACACACCTCGGCTAATGTTTGTATTTTATGTA 5460

QY 5461 GAGATGGGTTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCTCGACCTCAGAG 5514

Db 5461 GAGATGGGTTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCTCGACCTCAGAG 5514

RESULT 2

US-09-966-880A-35

; Sequence 35, Application US/09966880A

; Patent No. US2002016473A1

; GENERAL INFORMATION:

; APPLICANT: Honjo, Tasuku

; APPLICANT: Muramatsu, Masamichi

; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE

; FILE REFERENCE: 06501-088001

; CURRENT APPLICATION NUMBER: US/09/966, 880A

; PRIOR FILING DATE: 2001-09-28

; PRIOR APPLICATION NUMBER: PCT/JP00/01918

; PRIOR FILING DATE: 2000-03-28

; PRIOR FILING DATE: 1999-12-27

; PRIOR APPLICATION NUMBER: JP 11-371382

; PRIOR FILING DATE: 1999-12-27

; PRIOR APPLICATION NUMBER: JP 11-178999

; PRIOR FILING DATE: 1999-06-24

; PRIOR APPLICATION NUMBER: JP 11-87192

; PRIOR FILING DATE: 1999-03-29

; NUMBER OF SEQ ID NOS: 36

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 35

; LENGTH: 11204

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-966-880A-35

Query Match 86.5%; Score 4771; DB 9; Length 11204;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 4921; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 591 AGTTTCAGAGAGACTGTGGGAATATGGGGAATATAGAGGCTATCTGAGGCTCTTCAACAC 650

Db 1 AGTTTCAGAGAGACTGTGGGAATATGGGGAATATAGAGGCTATCTGAGGCTCTTCAACAC 60

QY 651 AATAACCCCAAGAGCTATTTAAATGCTCTTTAAAGTATTTTACATAAATATTACTATTCTC 710

Db 61 AATAACCCCAAGAGCTATTTAAATGCTCTTTAAAGTATTTTACATAAATATTACTATTCTC 120

QY 711 ATTGCTCTTTTATTTTGTGTATCATGATTATTAATGAAGTGTCTACTGTACTGCTCTCC 770

Db 121 ATTGCTCTTTTATTTTGTGTATCATGATTATTAATGAAGTGTCTACTGTACTGCTCTCC 180

QY 771 TGATCTTTGCTAGCTATGGAGCATGGAGTGGGCTTTTAGACAGCAGCAGCCCCAAAGGAACC 830

Db 181 TGATCTTTGCTAGCTATGGAGCATGGAGTGGGCTTTTAGACAGCAGCAGCCCCAAAGGAACC 240

QY 831 TAAACATTTAAAGCAGAGCTGCCCTCAATGGTTTAACTGTGTGACTCTGCTATGACAGC 890

Db 241 TAAACATTTAAAGCAGAGCTGCCCTCAATGGTTTAACTGTGTGACTCTGCTATGACAGC 300

QY 891 CCCACCCACCCTCTTCTACTTGGATCCAAATCAGGAGCAAGGCCGTTGGGGTACTCTGGTGG 950

Db 301 CCCACCCACCCTCTTCTACTTGGATCCAAATCAGGAGCAAGGCCGTTGGGGTACTCTGGTGG 360

QY 951 GGGTGTGTCTGTCTAGGGGAGGAGCCCAAAAGGGCAAGCTCAAAATTTGAATGTGAAGGCC 1010

Db 361 GGGTGTGTCTGTCTAGGGGAGGAGCCCAAAAGGGCAAGCTCAAAATTTGAATGTGAAGGCC 420

QY 1011 AATGCACTGTCTAGACTGTAGACAGAGAACCATCATTAATTTGAAGTGTAGATTTTCTGGCCT 1070

Db 421 AATGCACTGTCTAGACTGTAGACAGAGAACCATCATTAATTTGAAGTGTAGATTTTCTGGCCT 480

QY 1071 GAGACTTTCAGGGAGGCAAGAACACACTCTGGACACCACTATGGACAGGTTAAAGAGGCAG 1130

Db 481 GAGACTTTCAGGGAGGCAAGAACACACTCTGGACACCACTATGGACAGGTTAAAGAGGCAG 540

QY 1131 TCTTCTCGGGGTGATTGCACTGGCCCTTCTCTCAGAGCAAAATCTGAGTAATGAGACTGG 1190

Db 541 TCTTCTCGGGGTGATTGCACTGGCCCTTCTCTCAGAGCAAAATCTGAGTAATGAGACTGG 600


```
RESULT 3
US-09-966-880A-11
; Sequence 11, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 87
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-11

Query Match      1.6%; Score 87; DB 9; Length 87;
Best Local Similarity 100.0%; Pred. No. 8.7e-31;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1032 AGAACCATCATTAATTGAAGTGAAGTATTTCTGGCCTGAGACTTTCAGGAGGCAAGA 1091
Db 1 AGAACCATCATTAATTGAAGTGAAGTATTTCTGGCCTGAGACTTTCAGGAGGCAAGA 60

QY 1092 AGACACTCTGGACACCACTATGGACAG 1118
Db 61 AGACACTCTGGACACCACTATGGACAG 87

RESULT 4
US-09-966-880A-7
; Sequence 7, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 2818
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (80)...(673)
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(79)
; FEATURE:
```

```
; NAME/KEY: 3'UTR
; LOCATION: (677)...(2818)
US-09-966-880A-7

Query Match      1.6%; Score 87; DB 9; Length 2818;
Best Local Similarity 100.0%; Pred. No. 1.1e-30;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1032 AGAACCATCATTAATTGAAGTGAAGTATTTCTGGCCTGAGACTTTCAGGAGGCAAGA 1091
Db 1 AGAACCATCATTAATTGAAGTGAAGTATTTCTGGCCTGAGACTTTCAGGAGGCAAGA 60

QY 1092 AGACACTCTGGACACCACTATGGACAG 1118
Db 61 AGACACTCTGGACACCACTATGGACAG 87

RESULT 5
US-09-832-292-30
; Sequence 30, Application US/09832292
; Patent No. US20020177205A1
; GENERAL INFORMATION:
; APPLICANT: Ryazanov, Alexey
; TITLE OF INVENTION: MAMMALIAN ALPHA-KINASE PROTEINS, NUCLEIC ACIDS AND
; FILE REFERENCE: 601-1-098CIP
; CURRENT APPLICATION NUMBER: US/09/832,292
; CURRENT FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 09/632,131
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 8158
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-832-292-30

Query Match      1.2%; Score 68; DB 9; Length 8158;
Best Local Similarity 100.0%; Pred. No. 1.2e-21;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5446 TTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 5505
Db 6996 TTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 7055

QY 5506 ACCTCAGA 5513
Db 7056 ACCTCAGA 7063

RESULT 6
US-09-764-891-9880/c
; Sequence 9880, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9880
; LENGTH: 84
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-9880

Query Match      1.2%; Score 67; DB 9; Length 84;
Best Local Similarity 100.0%; Pred. No. 2.7e-21;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	5446	TTTGTA	TTTTT	ACTAG	ATGGG	TTTCA	CCATG	TTGCC	AGGCT	GGTCT	CAAACT	CCTG	5505
Dp	73	TTTGTA	TTTTT	ACTAG	ATGGG	TTTCA	CCATG	TTGCC	AGGCT	GGTCT	CAAACT	CCTG	14

```

RESULT 7
US-09-764-891-5684
; Sequence 5684, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 5684
; LENGTH: 1160
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-5684

```

```

RESULT 8
US-09-764-877-2384/c
; Sequence 2384, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2384
; LENGTH: 5930
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-2384

```

```

RESULT 9
US-09-984-827-5
; Sequence 5, Application US/09984827
; Publication No. US20030056234A1
; GENERAL INFORMATION:
; APPLICANT: DENEFFE, PATRICE
; APPLICANT: ROSIER-MONTUS, MARIE-FRANCOISE
; APPLICANT: ARNOULD-REGUIGNE, ISABELLE
; APPLICANT: DUVERGER, NICOLAS
; APPLICANT: CAMBIEN, FRANCOIS
; TITLE OF INVENTION: POLYMORPHIC SEQUENCES OF THE HUMAN ABCAL GENE, THEIR USE
; FILE REFERENCE: 03806.0522-00000
; CURRENT APPLICATION NUMBER: US/09/984,827
; CURRENT FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: 60/254,108
; PRIOR FILING DATE: 2000-12-11
; PRIOR APPLICATION NUMBER: FR 00/14037
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 161
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 11754
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-984-827-5

```

```

RESULT 10
; US-09-764-877-2385/c
; Sequence 2385, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 2385
; LENGTH: 16086
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-877-2385

```

Q7	5448	11409	5506	11349	1135
Db	TTTTATT	TTTGTATT	ACCTCAG	ACCTCAG	TTTGTATT
QY	TTTTATT	TTTGTATT	ACCTCAG	ACCTCAG	TTTGTATT
Db	TTTTATT	TTTGTATT	ACCTCAG	ACCTCAG	TTTGTATT


```
RESULT 11
US-09-764-891-8207/c
; Sequence 8207, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 8207
; LENGTH: 24028
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-8207

Query Match          1.2%  Score 67;  DB 9;  Length 24028;
Best Local Similarity 100.0%;  Pred. No. 3.7e-21;
Matches 67;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy  5446  TTTGTTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGTCTCAAACTCCTG 5505
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19797  TTTGTTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGTCTCAAACTCCTG 19738

Qy  5506  ACCTCAG 5512
      |||||||
Db  19737  ACCTCAG 19731

RESULT 12
US-10-074-095-1108/c
; Sequence 1108, Application US/10074095
; Publication No. US20030077704A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008C1
; CURRENT APPLICATION NUMBER: US/10/074,095
; CURRENT FILING DATE: 2002-02-14
; Prior Application Number: 09/764,860
; Prior Filing Date: 2001-01-17
; Prior Application Number: 60/179,065
; Prior Filing Date: 2000-01-31
; Prior Application Number: 60/180,628
; Prior Filing Date: 2000-02-04
; Prior Application Number: 60/214,886
; Prior Filing Date: 2000-06-28
; Prior Application Number: 60/217,487
; Prior Filing Date: 2000-07-11
; Prior Application Number: 60/225,758
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/220,963
; Prior Filing Date: 2000-07-26
; Prior Application Number: 60/217,496
; Prior Filing Date: 2000-07-11
; Prior Application Number: 60/225,447
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/218,290
; Prior Filing Date: 2000-07-14
; Prior Application Number: 60/225,757
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/226,868
; Prior Filing Date: 2000-08-22
; Prior Application Number: 60/216,647
; Prior Filing Date: 2000-07-07
; Prior Application Number: 60/225,267
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/216,880
; Prior Filing Date: 2000-07-07
; Prior Application Number: 60/225,270
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/251,869
; Prior Filing Date: 2000-12-08
; Prior Application Number: 60/235,834
; Prior Filing Date: 2000-09-27
; Prior Application Number: 60/234,274
; Prior Filing Date: 2000-09-21
; Prior Application Number: 60/234,223
; Prior Filing Date: 2000-09-21
; Prior Application Number: 60/228,924
; Prior Filing Date: 2000-08-30
; Prior Application Number: 60/224,518
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/236,369
; Prior Filing Date: 2000-09-29
; Prior Application Number: 60/224,519
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/220,964
; Prior Filing Date: 2000-07-26
; Prior Application Number: 60/241,809
; Prior Filing Date: 2000-10-20
; Prior Application Number: 60/249,299
; Prior Filing Date: 2000-11-17
; Prior Application Number: 60/236,327
; Prior Filing Date: 2000-09-29
; Prior Application Number: 60/241,785
; Prior Filing Date: 2000-10-20
; Prior Application Number: 60/244,617
; Prior Filing Date: 2000-11-01
; Prior Application Number: 60/225,268
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/236,368
; Prior Filing Date: 2000-09-29
; Prior Application Number: 60/251,856
; Prior Filing Date: 2000-12-08
; Prior Application Number: 60/251,868
; Prior Filing Date: 2000-12-08
; Prior Application Number: 60/229,344
; Prior Filing Date: 2000-09-01
; Prior Application Number: 60/234,997
; Prior Filing Date: 2000-09-25
; Prior Application Number: 60/229,343
; Prior Filing Date: 2000-09-01
; Prior Application Number: 60/229,345
; Prior Filing Date: 2000-09-01
; Prior Application Number: 60/229,287
; Prior Filing Date: 2000-09-01
; Prior Application Number: 60/229,513
; Prior Filing Date: 2000-09-05
; Prior Application Number: 60/231,413
; Prior Filing Date: 2000-09-08
; Prior Application Number: 60/229,509
; Prior Filing Date: 2000-09-05
; Prior Application Number: 60/236,367
; Prior Filing Date: 2000-09-29
; Prior Application Number: 60/237,039
; Prior Filing Date: 2000-10-02
; Prior Application Number: 60/237,038
; Prior Filing Date: 2000-10-02
; Prior Application Number: 60/236,370
; Prior Filing Date: 2000-09-29
; Prior Application Number: 60/236,802
; Prior Filing Date: 2000-10-02
; Prior Application Number: 60/237,037
; Prior Filing Date: 2000-10-02
; Prior Application Number: 60/237,040
; Prior Filing Date: 2000-10-02
; Prior Application Number: 60/240,960
; Prior Filing Date: 2000-10-20
; Prior Application Number: 60/239,935
; Prior Filing Date: 2000-10-13
; Prior Application Number: 60/239,937
; Prior Filing Date: 2000-10-13
```

;
; PRIOR APPLICATION NUMBER: 60/241,787
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,474
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,532
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,216
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,210
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/226,681
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,759
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/225,213
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/227,182
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,214
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/235,836
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/230,438
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/215,135
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/225,266
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,218
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,213
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,212
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,245
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,264
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399

;
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08

Query Match 1.2%; Score 67; DB 9; Length 32170;
Best Local Similarity 100.0%; Pred. No. 3.8e-21;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5446 TTTGTATTTTGTAGATGGGTTTTCACCATGTTGGCAGGCTGGTCTCAAACCTCCTG 5505
|||||
Db 28881 TTTGTATTTTGTAGATGGGTTTTCACCATGTTGGCAGGCTGGTCTCAAACCTCCTG 28822
|||||
QY 5506 ACCTCAG 5512
|||||
Db 28821 ACCTCAG 28815
|||||

RESULT 13
US-09-764-860-1108/c
; Sequence 1108, Application US/09764860
; Patent No. US20020094953A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008
; CURRENT APPLICATION NUMBER: US/09/764,860
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1108
; LENGTH: 32170
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-860-1108

Query Match 1.2%; Score 67; DB 10; Length 32170;
Best Local Similarity 100.0%; Pred. No. 3.8e-21;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5446 TTTGTATTTTGTAGATGGGTTTTCACCATGTTGGCAGGCTGGTCTCAAACCTCCTG 5505
|||||
Db 28881 TTTGTATTTTGTAGATGGGTTTTCACCATGTTGGCAGGCTGGTCTCAAACCTCCTG 28822
|||||
QY 5506 ACCTCAG 5512
|||||
Db 28821 ACCTCAG 28815
|||||

RESULT 14
US-09-900-449A-3/c
; Sequence 3, Application US/09900449A
; Publication No. US20030040616A1
; GENERAL INFORMATION:
; APPLICANT: ZHONG, Jenny et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; FILE REFERENCE: CL001271
; CURRENT APPLICATION NUMBER: US/09/900,449A
; CURRENT FILING DATE: 2001-07-09

; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 34668
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-900-449A-3

Query Match 1.2%; Score 67; DB 9; Length 34668;
Best Local Similarity 100.0%; Pred. No. 3.8e-21;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5446 TTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGTCTCAAACTCCTG 5505
Db 23213 TTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGTCTCAAACTCCTG 23154
QY 5506 ACCTCAG 5512
Db 23153 ACCTCAG 23147

RESULT 15

US-09-982-091A-5
; Sequence 5, Application US/09982091A
; Patent No. US20020151030A1
; GENERAL INFORMATION:
; APPLICANT: CALIFORNIA INSTITUTE OF TECHNOLOGY
; APPLICANT: KUMAGAI, Akiko
; APPLICANT: DUNPHY, William
; TITLE OF INVENTION: CLASPIN PROTEINS AND METHODS OF USE THEREOF
; FILE REFERENCE: CIT1320-1
; CURRENT APPLICATION NUMBER: US/09/982,091A
; CURRENT FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: US 60/241,246
; PRIOR FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 58837
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-982-091A-5

Query Match 1.2%; Score 67; DB 10; Length 58837;
Best Local Similarity 100.0%; Pred. No. 3.9e-21;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5446 TTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGTCTCAAACTCCTG 5505
Db 55194 TTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGTCTCAAACTCCTG 55253
QY 5506 ACCTCAG 5512
Db 55254 ACCTCAG 55260

Search completed: June 19, 2003, 17:54:07
Job time : 722 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 18:30:49 ; Search time 221.5 Seconds
(without alignments)
6634.257 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 1736
Sequence: 1 acagacgaatacatagctgtcca.....tcaaactcctgacctcagag 5514

Scoring table:
OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Word size: 1
Total number of hits satisfying chosen parameters: 1687582

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Command line parameters:
-MODE=frame+n2p.model -DEV=xlp
-Q=/cyn2_1/USPto_spool/US09966880/runat_14062003_175620_11018/app_query.fasta_1.5703
-DB=A_Geneseq_101002 -QFMT=fastan -SUFFIX=olin2p.rag -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -NATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09966880 -CGCN_1_1_379 @runat_14062003_175620_11018 -NCPU=6 -ICPU=3
-NO_WMAP -LARGESQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002.*
1: /SID32/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID32/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID32/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID32/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID32/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID32/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID32/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID32/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID32/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID32/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID32/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID32/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID32/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID32/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID32/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID32/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID32/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID32/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID32/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID32/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID32/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID32/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID32/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	39	2.2	251	22	ABG17006
c 2	22	1.3	75	22	AAO02240
c 3	22	1.3	103	21	AAB43997
c 4	22	1.3	110	21	AAB43292
c 5	22	1.3	116	22	AAB93273
c 6	22	1.3	137	22	ABG19377
c 7	22	1.3	407	22	AAU30450
c 8	21	1.2	80	22	AAU86859
c 9	21	1.2	118	22	AAU93908
c 10	20	1.2	36	22	AAO12441
c 11	20	1.2	42	22	ABG11938
c 12	20	1.2	89	22	ABG09800
c 13	20	1.2	125	22	AAU41856
c 14	19	1.1	41	22	AAO05250
c 15	19	1.1	55	22	AAO09035
c 16	19	1.1	93	22	AAU32720
c 17	19	1.1	110	23	ABP41552
c 18	19	1.1	115	22	AAU96287
c 19	19	1.1	115	22	AAU42316
c 20	19	1.1	115	23	ABG68710
c 21	19	1.1	136	23	AAE14727
c 22	19	1.1	141	20	AAU19767
c 23	19	1.1	188	22	ABG11932
c 24	18	1.0	44	22	AAO05621
c 25	18	1.0	60	22	AAU86602
c 26	18	1.0	63	22	AAO02094
c 27	18	1.0	81	22	AAU96266
c 28	18	1.0	81	22	AAU42307
c 29	18	1.0	86	21	AAU42412
c 30	18	1.0	90	22	AAU30185
c 31	18	1.0	99	22	ABG78806
c 32	18	1.0	104	22	AAO01382
c 33	18	1.0	106	22	AAU08280
c 34	18	1.0	131	22	AAU29521
c 35	17	1.0	25	22	ABG08228
c 36	17	1.0	33	22	AAO08563
c 37	17	1.0	34	22	AAU31173
c 38	17	1.0	35	22	AAO11772
c 39	17	1.0	37	22	AAO05588
c 40	17	1.0	37	22	AAO13714
c 41	17	1.0	38	22	AAO13748
c 42	17	1.0	39	22	AAO03987
c 43	17	1.0	40	22	AAO07106
c 44	17	1.0	43	22	AAU62290
c 45	17	1.0	45	22	ABG05119

ALIGNMENTS

RESULT 1
ABG17006
ID ABG17006 standard; Protein; 251 AA.
XX
AC ABG17006;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #16997.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX

PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PR
PA (HYSE-) HYSEQ INC.
XX
XX

Drmanac RT, Liu C, Tang YT;

WPI: 2001-639362/73.
N-PSDB; AAS81193.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -

Claim 20; SEQ ID No 47365; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG3037 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 251 AA;

Alignment Scores:

Pred. No.:	3,1e-28	Length:	251
Score:	39.00	Matches:	39
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.24%	Indels:	0
DB:	22	Gaps:	0

US-09-966-880A-9 (1-5514) x ABG17006 (1-251)

QY	1017	GTGCATTGGCCCTCACATTCAAAATTGAGCTTGCCCTTTGGGCTCTCCCTGACACG	958
Db	169	ValHisTrpProPheThrPhelysPheGluLeuAlaLeuLeuGlySerSerProAspSer	188
QY	957	ATCACCCACACAGGTACCCCAACGGCCTTGCTCTGATTGGATCCAGTGAAGATG	901
Db	189	IleThrProThrArgTyrProAsnGlyLeuAlaProAspLeuAspProVallylMet	207

RESULT 2

AAO02240
ID AAO02240 standard; Protein; 75 AA.

XX AAO02240;

XX 06-NOV-2001 (first entry)

XX Human polypeptide SEQ ID NO 16132.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

PN WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

WPI: 2001-514838/56.

DR N-PSDB; AAI82171.

XX Isolated nucleic acids and polypeptides, useful for preventing

PT diagnosing and treating e.g. leukaemia, inflammation and immune

PT disorders -

XX Claim 20; SEQ ID NO 16132; 1399pp + Sequence Listing; English.

CC The invention relates to human polynucleotides (AAI79941-AAI93841) and the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 75 AA;

Alignment Scores:

Pred. No.:	9,03e-12	Length:	75
Score:	22.00	Matches:	22
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.27%	Indels:	0
DB:	22	Gaps:	0

US-09-966-880A-9 (1-5514) x AAO02240 (1-75)

QY	3623	TCAAGCGATCCACCCACCTCAGCCTCCAGAGTGCTGGGATTACAGCGGTGAGCCACCAC	3564
Db	35	SerSerAspProProThrSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHis	54

QY 3563 ACCTGG 3558

Db 55 ThrTrp 56

RESULT 3

AAAB43997

ID AAB43997 standard; Protein; 103 AA.

XX AAB43997;

XX 08-FEB-2001 (first entry)

XX Human cancer associated protein sequence SEQ ID NO:1442.

KW Human; cancer associated gene; cancer antigen; detection; cancer;

KW diagnosis; cytostatic; proliferative; vulnery; immunomodulator;
 KW antidiabetic; antiaesthatic; antirheumatic; antiarthritic; antiviral;
 KW antinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
 KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
 KW vasotropic; antipsoriatic; angiogenic; gene therapy; inflammation;
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
 KW allergic reaction; graft versus host disease; organ rejection;
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;
 KW neurological disease; drug screening.
 XX
 OS Homo sapiens.
 XX
 PN WO200055350-A1.
 XX
 XX 21-SEP-2000.
 XX
 XX 08-MAR-2000; 2000WO-US05882.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2000-587533/55.
 DR N-PSDB; AAC78206.
 XX
 PT Novel isolated nucleic acids comprising sequences encoding peptides
 PT useful for treating or diagnosing e.g. cancer -
 XX
 XX Claim 11; Page 2126; 2352pp; English.
 PS
 CC AAC77607 to AAC78448 encode the human cancer associated proteins given
 CC in AAB43398 to AAB44239. The proteins can have activities based on the
 CC tissues and cells the genes are expressed in. Example of activities
 CC include: cytostatic; proliferative; vulnery; immunomodulator;
 CC antidiabetic; antiaesthatic; antirheumatic; antiarthritic;
 CC antinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
 CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
 CC nootropic; vasotropic; antipsoriatic and angiogenic. The
 CC polynucleotides and polypeptides can be used for preventing, treating or
 CC ameliorating medical conditions and diagnosing pathological conditions.
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
 CC the present invention may be used to treat immune disorders by activating
 CC or inhibiting the proliferation, differentiation or mobilisation of
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune
 CC disorders, allergic reactions, graft versus host disease and organ
 CC rejection, modulate haemostatic or thrombolytic activity, modulate
 CC inflammation, cancers, cardiovascular disorders, neurological disease and
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of
 CC the present invention.
 XX
 SQ Sequence 103 AA;
 Alignment Scores:
 Pred. No.: 8,64e-12 Length: 103
 Score: 22.00 Matches: 22
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.27% Indels: 0
 DB: 21 Gaps: 0
 US-09-966-880A-9 (1-5514) x AAB43997 (1-103)
 QY 5448 TGTAFTTTTAGTAGAGTGGGTTTCACCATGTTGCCAGGCTGCTCAAACTCCTGAC 5507
 |||||||
 Db 82 CysllePheSerArgAspGlyValSerProCysTrpProGlyTrpSerGlnThrProAsp 101
 QY 5508 CTCAGA 5513
 |||||||
 Db 102 LeuArg 103

RESULT 4

AAB43292

ID AAB43292 standard; Protein; 110 AA.

XX AAB43292;

AC AAB43292;

XX 08-FEB-2001 (first entry)

XX Human ORFX ORF3056 polypeptide sequence SEQ ID NO:6112.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 XX Shimkets RA, Leach M;
 PI WPI; 2000-602362/57.
 XX N-PSDB; AAC77501.
 DR Novel nucleic acids and peptides derived from open reading frame X,
 DR useful for treating e.g. cancers, proliferative disorders,
 DR neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 11; Page 5297-5298; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX SQ Sequence 110 AA;

Alignment Scores: Length: 110
Pred. No.: 8.56e-12
Score: 22.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.27%
DB: 21

US-09-966-880A-9 (1-5514) x AAB43292 (1-110)

QY 5448 TGTATTATTAGAGATGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCTGAC 5507
DB 62 CysIlePheSerArgaspGlyValSerProCysTrpProGlyTrpSerGlnThrProAsp 81

QY 5508 CTCAGA 5513
DB 82 LeuArg 83

RESULT 5
AAB93273
ID AAB93273 standard; Protein; 116 AA.
AC AAB93273;
DT 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:12312.
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.
KW Homo sapiens.
OS Homo sapiens.
PN EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX Claim 8; SEQ ID 12312; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAB03166 to AAB13628 and
CC AAB13633 to AAB18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAB13629 to AAB13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX

SQ Sequence 116 AA;

Alignment Scores: Length: 116
Pred. No.: 8.5e-12
Score: 22.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.27%
DB: 22

US-09-966-880A-9 (1-5514) x AAB93273 (1-116)

QY 5446 TTGTATTATTAGAGATGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCTG 5505
DB 51 PheValPheLeuValGluMetGlyPheHisHisValGlyGlnAlaGlyLeuLysLeuLeu 70

QY 5506 ACCTCA 5511
DB 71 ThrSer 72

RESULT 6
ABG19377
ID ABG19377 standard; Protein; 137 AA.
XX AC ABG19377;
XX 18-FEB-2002 (first entry)
DT Novel human diagnostic protein #19368.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
KW Homo sapiens.
OS Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX N-PSDB; AAS83564.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID NO 49736; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 137 AA;

Alignment Scores: 8.31e-12 Length: 137
 Pred. No.: 22.00 Matches: 22
 Score: 100.00% Conservative: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 1.27% Indels: 0
 Query Match: 22 Gaps: 0
 DB:

US-09-966-880A-9 (1-5514) x ABG19377 (1-137)

QY 3559 CAGGTGGTGGCTCACCGCTGTAAATCCAGCACTCTGGAGGTGGTGGAFCG 3618
 Db 64 GlnValTrpTrpLeuThrProValIleProAlaLeuTrpGluAlaGluValIcGlySer 83
 QY 3619 CTTGAG 3624
 Db 84 LeuGlu 85

RESULT 7
 AAU30450
 ID AAU30450 standard; Protein: 407 AA.
 XX
 AC AAU30450;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human secreted protein #941.

XX Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200179449-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 16-APR-2001; 2001WO-US08656.
 XX
 PR 18-APR-2000; 2000US-0552929.
 PR 26-JAN-2001; 2001US-0770160.
 XX
 PA (HYSEQ-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-611725/70.
 XX
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy -
 XX
 PS Claim 20; Page 294; 765pp; English.

XX The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising
 CC the nucleic acids encoding the polypeptides and cells genetically
 CC engineered to express them are also useful for producing the proteins.
 CC The proteins are useful in genetic vaccination, testing and
 CC therapy, and can be used as nutritional supplements. They may be used to
 CC increase stem cell proliferation; to regulate haematopoiesis; and in
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and
 CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
 CC sequences of novel human secreted proteins of the invention.
 XX
 SQ Sequence 407 AA;

Alignment Scores: 7.15e-12 Length: 407
 Pred. No.: 22.00 Matches: 22
 Score: 100.00% Conservative: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 1.27% Indels: 0
 Query Match: 22 Gaps: 0
 DB:

US-09-966-880A-9 (1-5514) x AAU30450 (1-407)

QY 5446 TTTGTATTTTAGAGATGGGTTTCCACCATGTTGCCAGGCTGTCAACTCCTG 5505
 Db 63 PheValPheLeuValGluMetGlyPheHisValGlyGlnAlaGlyLeuLysLeuLeu 82
 QY 5506 ACCTCA 5511
 Db 83 ThrSer 84

RESULT 8
 AAU86859
 ID AAU86859 standard; Protein: 80 AA.
 XX
 AC AAU86859;
 XX
 DT 07-NOV-2001 (first entry)
 XX
 DE Human immune/haematopoietic antigen SEQ ID NO:14452.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis.
 XX
 OS Homo sapiens.
 XX
 PN WO200157182-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234597.
 PR 25-SEP-2000; 2000US-0234598.
 PR 25-SEP-2000; 2000US-0234584.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-483426/52.

DR N-PSDB; AAK59640.

PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -

PS Claim 11; SEQ ID NO 14452; 3071pp + Sequence Listing; English.

XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)

CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AA64703
CC to AA87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AA834942 to AA854950 and AA882169
CC represent sequences used in the exemplification of the present invention.

XX SQ Sequence 80 AA;

Alignment Scores: 8.24e-11 Length: 80
Pred. No.: 21.00 Matches: 21
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 1.21% Gaps: 0
DB: 22

US-09-966-880A-9 (1-5514) x AA86859 (1-80)

QY 3562 GTGTGGTCTACGGCTGTAAATCCAGACTCTGGAGGCTGAGTGGTGATCGCTT 3621

Db 43 ValTrpIrpLeuThrProValIleProAlaLeuTrpGluAlaGluValGlySerLeu 62
|||||

QY 3622 GAG 3624

Db 63 Glu 63
|||

RESULT 9
AAB93908

ID AAB93908 standard; Protein: 118 AA.

XX AC AAB93908;

XX 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:13871.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -

XX Claim 8; SEQ ID 13871; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX SQ Sequence 118 AA;

Alignment Scores: 7.81e-11 Length: 118
Pred. No.: 21.00 Matches: 21
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 1.21% Gaps: 0
DB: 22

US-09-966-880A-9 (1-5514) x AAB93908 (1-118)

QY 5448 TGTATTTTATGATGAGATGGGTTTCACCATGTTGGCAGGCTGTCTCAAACTCCTGAC 5507
|||||

Db 51 CysIlePheSerArgAspGlyValSerProCysTrpProGlyTrpSerGlnThrProAsp 70

QY 5508 CTC 5510

Db 71 Leu 71
|||

RESULT 10
AAO12441

ID AAO12441 standard; Protein: 36 AA.

XX AC AAO12441;

XX 06-NOV-2001 (first entry)

XX Human polypeptide SEQ ID NO 26333.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

XX WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

XX 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

XX N-PSDB; AA192372.

XX Isolated nucleic acids and polypeptides, useful for preventing

PT diagnosing and treating e.g. leukaemia, inflammation and immune disorders -

XX Claim 20; SEQ ID NO 26333; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AA179941-AA193841) and the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 36 AA;

Alignment Scores:

Pred. No.:	8.47e-10	Length:	36
Score:	20.00	Matches:	20
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.15%	Indels:	0
DB:	22	Gaps:	0

US-09-966-880A-9 (1-5514) x AAO12441 (1-36)

QY 3565 TGGTGGCTCACGCTCTAATCCAGCACCTGGAGGCTGAGTGGTGGCTTGAG 3624
ABG11938

DB 15 TrpTrpLeuThrProValIleProAlaLeuTrpGluAlaGluValGlyGlySerLeuGlu 34

RESULT 11

ID ABG11938 standard; Protein; 42 AA.

AC ABG11938;

XX 18-FEB-2002 (first entry)

DT

DE Novel human diagnostic protein #11929.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

KW

OS Homo sapiens.

XX WO200175067-A2.

PN

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

PF

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

PR

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

PI

XX WPI; 2001-639362/73.

XX N-PSDB; AAS76125.

DR

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess

PT biodiversity -

PT

XX Claim 20; SEQ ID No 42297; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABO00010-ABG030377 represent novel human diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 42 AA;

Alignment Scores:

Pred. No.:	8.3e-10	Length:	42
Score:	20.00	Matches:	20
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.15%	Indels:	0
DB:	22	Gaps:	0

US-09-966-880A-9 (1-5514) x ABG11938 (1-42)

QY 5452 TTTTGTAGTAGAGTGGGTTTCACCATGTTGGCAGGCTGGTCTCAACTCCTGACCTCA 5511
|||||

DB 5 PheLeuValGluMetGlyPheHisHisValGlyGlnAlaGlyLeuLysLeuThrSer 24

RESULT 12

ABG09800

ID ABG09800 standard; Protein; 89 AA.

AC ABG09800;

XX 13-FEB-2002 (first entry)

DT

DE Novel human diagnostic protein #9791.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

KW

OS Homo sapiens.

XX WO200175067-A2.

PN

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

PF

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

PR

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

PI

XX WPI; 2001-639362/73.

XX N-PSDB; AAS73987.

DR

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess

PT biodiversity -

PT

XX Claim 20; SEQ ID No 42297; 103pp; English.

PT biodiversity -
PS Claim 20; SEQ ID NO 40159; 103pp; English.
XX
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABC30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 89 AA;

Alignment Scores:
Pred. No.: 7.48e-10 Length: 89
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.15% Indels: 0
DB: 22 Gaps: 0

US-09-966-880A-9 (1-5514) x ABG09800 (1-89)
QY 5454 TTTAGTAGAGATGGGTTTACCATGTTGGCCAGGCTGCTCAAACTCCTGACCTCAGA 5513
Db 70 PheSerArgSpGlyValSerProCysIrpProGlyIrpSerGlnThrProAspIueArg 89

RESULT 13
AA041856
ID AA041856 standard; Protein; 125 AA.
XX
AC AA041856;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 6787.
XX
KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0596042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
DR N-PSDB; AA161012.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 6787; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AA038642-AA042213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX assays for receptor activity, cancer diagnosis and therapy, drug screening,
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
SQ Sequence 125 AA;

Alignment Scores:
Pred. No.: 7.14e-10 Length: 125
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.15% Indels: 0
DB: 22 Gaps: 0

US-09-966-880A-9 (1-5514) x AA041856 (1-125)
QY 5449 GTATTTTAGTAGAGATGGGTTTACCATGTTGGCCAGGCTGCTCAAACTCCTGACC 5508
Db 90 ValPheLeuValGluMetGlyPheHisValGlyGlnAlaGlyLeuLysLeuLeuThr 109

RESULT 14
AA05250
ID AA05250 standard; Protein; 41 AA.
XX
XX AA05250;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 19142.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
XX Homo sapiens.
XX
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US04927.
XX

```
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Drmanac RT;
PI WPI: 2001-514838/56.
XX N-PSDB; AA185181.
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders.
XX Claim 20; SEQ ID NO 19142; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and/or
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 41 AA;
Alignment Scores:
Pred. No.: 7.67e-09 Length: 41
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 22 Gaps: 0
US-09-966-880A-9 (1-5514) x AAO05250 (1-41)
QY 3614 CCACCACCTCAGCTCCAGAGTGTGGATTACAGGCGTGAGCCACACACCTGG 3558
Db 21 ProProThrSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHistrTrp 39
RESULT 15
AAO09035
ID AAO09035 standard; Protein; 55 AA.
XX AC AAO09035;
XX 06-NOV-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 22927.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX OS Homo sapiens.
XX PN WO200164835-A2.
XX PD 07-SEP-2001.
XX PF 26-FEB-2001; 2001WO-US04927.
XX 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
```

```
PA (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
XX WPI: 2001-514838/56.
XX N-PSDB; AA188956.
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders.
XX Claim 20; SEQ ID NO 22927; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and/or
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 55 AA;
Alignment Scores:
Pred. No.: 7.36e-09 Length: 55
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 22 Gaps: 0
US-09-966-880A-9 (1-5514) x AAO09035 (1-55)
QY 3303 GTCTTGAACCTCCTGGGCTCAAGCAATCCTCGCTGGCCTCCCAAGTGTGGGA 3359
Db 34 ValLeuAsnSerTrpAlaGlnAlaLeuLeuLeuProTrpProPolysValLeuGly 52
Search completed: June 14, 2003, 19:16:33
Job time : 234.5 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 14, 2003, 19:16:40 ; Search time 174 Seconds
(without alignments)
6776.242 Million cell updates/sec

Title: US-09-966-880A-9
Perfect score: 1736
Sequence: 1 acagacgaatacatagtccca.....tcaaaactcctgacctcagag 5514

Scoring table:
OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 408643 seqs, 106915682 residues

Word size: 1

Total number of hits satisfying chosen parameters: 764838

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-Q/cgn2_1/USPTO_spool/US09966880/runat_14062003_175623_11119/app_query.fasta_1.5703
-DB=Published.Applications_AA -OFMT=fastan -SUFFIX=olin2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09966880@cgn.1.1.63@runat_14062003_175623_11119
-NCPU=6 -ICPU=3 -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=6 -DELOP=6 -DELEXT=7

Database : Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PTCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PTCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	1.3	103	10	US-09-925-301-1442
2	19	1.1	115	9	US-09-764-891-4945
3	19	1.1	115	10	US-09-995-494-79
4	18	1.0	81	9	US-09-764-891-4924

Sequence	1350, Ap	Sequence 111, App	Sequence 3289, Ap	Sequence 712, App	Sequence 115, App	Sequence 207, App	Sequence 833, App	Sequence 742, App	Sequence 4886, Ap	Sequence 150, App	Sequence 187, App	Sequence 280, App	Sequence 105, App	Sequence 869, App	Sequence 33, Appl	Sequence 282, App	Sequence 433, App	Sequence 140, App	Sequence 1013, Ap	Sequence 1013, Ap	Sequence 1868, Ap	Sequence 829, App	Sequence 4372, Ap	Sequence 34, Appl	Sequence 454, App	Sequence 207, App	Sequence 202, App	Sequence 3138, Ap	Sequence 85, Appl	Sequence 1350, Ap	Sequence 3498, Ap	Sequence 1649, Ap	Sequence 288, App	Sequence 129, App	Sequence 464, App	Sequence 464, App	Sequence 1026, Ap	Sequence 1026, Ap	Sequence 352, App	Sequence 171, App	
Sequence 1350, Ap	17	17	17	17	17	17	16	16	16	16	16	16	16	16	16	16	16	16	15	15	15	15	15	15	15	15	15	14	14	14	14	14	14	14	14	14	14	14	14	14	
Sequence 111, App	6	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Sequence 3289, Ap	17	17	17	17	17	17	16	16	16	16	16	16	16	16	16	16	16	16	15	15	15	15	15	15	15	15	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
Sequence 712, App	6	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Sequence 115, App	17	17	17	17	17	17	16	16	16	16	16	16	16	16	16	16	16	16	15	15	15	15	15	15	15	15	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
Sequence 207, App	6	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Sequence 833, App	17	17	17	17	17	17	16	16	16	16	16	16	16	16	16	16	16	16	15	15	15	15	15	15	15	15	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
Sequence 742, App	6	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Sequence 4886, Ap	17	17	17	17	17	17	16	16	16	16	16	16	16	16	16	16	16	16	15	15	15	15	15	15	15	15	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
Sequence 150, App	6	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Sequence 187, App	17	17	17	17	17	17	16	16	16	16	16	16	16	16	16	16	16	16	15	15	15	15	15	15	15	15	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
Sequence 280, App	6	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Sequence 105, App	17	17	17	17	17	17	16	16	16	16	16	16	16	16	16	16	16	16	15	15	15	15	15	15	15	15	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
Sequence 869, App	6	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Sequence 33, Appl	17	17	17	17	17	17	16	16	16	16	16	16	16	16	16	16	16	16	15	15	15	15	15	15	15	15	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
Sequence 282, App	6	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Sequence 433, App	17	17	17	17	17	17	16	16	16	16	16	16	16	16	16	16	16	16	15	15	15	15	15	15	15	15	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
Sequence 140, App	6	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Sequence 1013, Ap	17	17	17	17	17	17	16	16	16	16	16	16	16	16	16	16	16	16	15	15	15	15	15	15	15	15	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
Sequence 1013, Ap	6	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Sequence 1868, Ap	17	17	17	17	17	17	16	16	16	16	16	16	16	16	16	16	16	16	15	15	15	15	15	15	15	15	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
Sequence 829, App	6	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Sequence 4372, Ap	17	17	17	17	17	17	16	16	16	16	16	16	16	16	16	16	16	16	15	15	15	15	15	15	15	15	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
Sequence 34, Appl	6	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Sequence 454, App	17	17	17	17	17	17	16	16	16	16	16	16	16	16	16	16	16	16	15	15	15	15	15	15	15	15	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
Sequence 207, App	6	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Sequence 202, App	17	17	17	17	17	17	16	16	16	16	16	16	16	16	16	16	16	16	15	15	15	15	15	15	15	15	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
Sequence 3138, Ap	6	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Sequence 85, Appl	17	17	17	17	17	17	16	16	16	16	16	16	16	16	16	16	16	16	15	15	15	15	15	15	15	15	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
Sequence 1350, Ap	6	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Sequence 3498, Ap	17	17	17	17	17	17	16	16	16	16	16	16	16	16	16	16	16	16	15	15	15	15	15	15	15	15	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
Sequence 1649, Ap	6	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Sequence 288, App	17	17	17	17	17	17	16	16	16	16	16	16	16	16	16	16	16	16	15	15	15	15	15	15	15	15	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
Sequence 129, App	6	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Sequence 464, App</																																									

```
Alignment Scores:
Pred. No.:      1.22e-10      Length:      103
Score:          22.00         Matches:      22
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    1.27%        Indels:      0
DB:             10           Gaps:      0

US-09-966-880A-9 (1-5514) x US-09-925-301-1442 (1-103)
QY 5448 TGTATTTTAGTAGAGATGGGTTTCACCATGTTGGCAGGCTGGTCTCAAACTCCTGCAC 5507
Db 82 CysIlePheSerArgAspGlyValSerProCysTrpProGlyTrpSerGlnThrProAsp 101
QY 5508 CTCAGA 5513
Db 102 LeuArg 103

RESULT 2
US-09-966-891-4945
; Sequence 4945, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4945
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (13)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (57)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (85)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (86)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (87)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (94)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (95)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (96)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (102)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (110)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (111)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
```

```
; NAME/KEY: SITE
; LOCATION: (113)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-4945

Alignment Scores:
Pred. No.:      5.89e-08      Length:      115
Score:          19.00         Matches:      19
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    1.09%        Indels:      0
DB:             9           Gaps:      0

US-09-966-880A-9 (1-5514) x US-09-764-891-4945 (1-115)
QY 5446 TTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGGCAGGCTGGTCTCAAACTC 5502
Db 38 PheValPheLeuValGluMetGlyPheHisHisValGlyGlnAlaGlyLeuLysLeu 56

RESULT 3
US-09-995-494-79
; Sequence 79, Application US/09995494
; Patent No. US20020127578A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Cafferkey, Robert
; APPLICANT: Ali, Shujath
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; APPLICANT: Chen, Sei-Yu
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes a
; FILE REFERENCE: DEX-0293
; CURRENT APPLICATION NUMBER: US/09/995,494
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 60/253,176
; PRIOR FILING DATE: 2000-11-27
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 79
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-995-494-79

Alignment Scores:
Pred. No.:      5.89e-08      Length:      115
Score:          19.00         Matches:      19
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    1.09%        Indels:      0
DB:             10           Gaps:      0

US-09-966-880A-9 (1-5514) x US-09-995-494-79 (1-115)
QY 3614 CCACCCACCTCAGCTCCAGAGTCTGGGATTACAGGCTGAGCCACACCTGG 3558
Db 95 ProProThrSerAlaSerGlnSerAlaGlylleThrGlyValSerHisHisThrtrp 113

RESULT 4
US-09-764-891-4924
; Sequence 4924, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
```


; SEQ ID NO 4924
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-891-4924

Alignment Scores:
Pred. No.: 4.92e-07 Length: 81
Score: 18.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.04% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-764-891-4924 (1-81)

QY 3574 ACGCTGTAAATCCAGCACTCTGGAGGCTGAGGTGGTGGATCGCTTGAGCCC 3627
Db 38 ThrProValIleProAlaLeuTrpGluAlaGluValGlyGlySerLeuGluPro 55
|||||

RESULT 5

US-09-764-877-1350
; Sequence 1350, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1350
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-877-1350

Alignment Scores:
Pred. No.: 4.07e-06 Length: 60
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.98% Indels: 0
DB: 10 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-764-877-1350 (1-60)

QY 5446 TTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGTCTC 5496
Db 10 PheValPheLeuValGluMetGlyPheHisValGlyGlnAlaGlyLeu 26
|||||

RESULT 6

US-09-820-843A-111
; Sequence 111, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 111
; LENGTH: 71
; TYPE: PRT
; ORGANISM: L. major
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: hypothetical protein Fl105.01

; NAME/KEY: misc_feature
; OTHER INFORMATION: gi16996498
US-09-820-843A-111

Alignment Scores:
Pred. No.: 3.96e-06 Length: 71
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.98% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-820-843A-111 (1-71)

QY 2890 TGTGTTTGT 2840
Db 40 CysValCysValCysValCysValCysValCysValCysValCysValCysValCys 56
|||||

RESULT 7

US-09-764-891-3289
; Sequence 3289, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3289
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (82)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-3289

Alignment Scores:
Pred. No.: 3.76e-06 Length: 98
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.98% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-764-891-3289 (1-98)

QY 5446 TTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGGTCTC 5496
Db 16 PheValPheLeuValGluMetGlyPheHisValGlyGlnAlaGlyLeu 32
|||||

RESULT 8

US-09-908-711-115
; Sequence 115, Application US/09908711
; Patent No. US20020045230A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA128
; CURRENT APPLICATION NUMBER: US/09/908,711
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US01/01360
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,867
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01344
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,892
; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: US01/01345
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,888
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01329
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,905
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01354
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,891
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01339
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,869
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01340
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,874
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01334
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,898
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01320
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,853
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01349
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,902
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01239
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,870
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01348
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,882
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01347
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,896
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01307
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,864
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01341
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,856
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01336
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,868
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01312
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 115
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (82)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-908-711-115

Alignment Scores:
Pred. No.: 3.76e-06 Length: 98
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.98% Indels: 0
DB: 10 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-908-711-115 (1-98)

QY 5446 TTTGTATTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTC 5496
|||||
Db 16 PheValPheLeuValGluMetGlyPheHisHisValGlyGlnAlaGlyLeu 32

RESULT 9

US-10-082-830-207
; Sequence 207, Application US/10082830
; Publication No. US20030077604A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Salceda, Susana
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; TITLE OF INVENTION: Genes and Proteins
; FILE REFERENCE: DEX-0249
; CURRENT APPLICATION NUMBER: US/10/082,830
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,802
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 207
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-082-830-207

Alignment Scores:
Pred. No.: 3.61e-06 Length: 126
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.98% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-9 (1-5514) x US-10-082-830-207 (1-126)

QY 5446 TTTGTATTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTC 5496
|||||
Db 83 PheValPheLeuValGluMetGlyPheHisHisValGlyGlnAlaGlyLeu 99

RESULT 10

US-09-925-297-833
; Sequence 833, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR FILING DATE: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 833


```

; SEQ ID NO 187
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-016-157A-187

Alignment Scores:
Pred. No.:      2.96e-05      Length:      100
Score:          16.00         Matches:     16
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:      0.92%      Indels:       0
DB:               9          Gaps:         0

US-09-966-880A-9 (1-5514) x US-10-016-157A-187 (1-100)

QY 5446 TTGTATTATTTAGTAGAGATGGGTTTCACCATGTTGGCCAGGCTGGT 5493
Db 57 PheValPheLeuValGluMetGlyPheHisHisValGlyGlnAlaGly 72

RESULT 15
US-10-074-475-280
; Sequence 280, Application US/10074475
; Publication No. US20030092898A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Hu, Ping
; APPLICANT: Recipon, Herve
; APPLICANT: Kariä, Kaipana
; APPLICANT: Cafferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; FILE REFERENCE: DEX-0313
; CURRENT APPLICATION NUMBER: US/10/074,475
; CURRENT FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: 60/268,292
; PRIOR FILING DATE: 2001-02-13
; NUMBER OF SEQ ID NOS: 295
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 280
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-074-475-280

Alignment Scores:
Pred. No.:      2.95e-05      Length:      101
Score:          16.00         Matches:     16
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:      0.92%      Indels:       0
DB:               9          Gaps:         0

US-09-966-880A-9 (1-5514) x US-10-074-475-280 (1-101)

QY 3614 CCACCCACCTCAGCTCCAGAGTGTGGATTACAGGCTGAGCCAC 3567
Db 86 ProProThrSerAlaSerGlnSerAlaGlyIleThrGlyValSerHis 101

Search completed: June 14, 2003, 19:43:11
Job time : 183 secs

```